ISSN 1684–8853

ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ

НАУЧНЫЙ ЖУРНАЛ

3(64)/2013

НАУЧНЫЙ ЖУРНАЛ

3(64)/2013

РЕЦЕНЗИРУЕМОЕ ИЗДАНИЕ

ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ

Адрес редакции: 190000, Санкт-Петербург,
Б. Морская ул., д. 67, ГУАП, РИЦ
Тел.: (812) 494-70-02, e-mail: 80x@mail.ru, сайт: www.i-us.ru

Журнал зарегистрирован в Министерстве РФ по делам печати, телерадиовещания и средств массовых коммуникаций. Свидетельство о регистрации ПИ № 77-12412 от 19 апреля 2002 г. Перерегистрирован в Роскомнадзоре. Свидетельство о регистрации ПИ № ФС77-49181 от 30 марта 2012 г.

Журнал входит в «Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени доктора и кандидата наук».

Журнал распространяется по подписке. Подписку можно оформить через редакцию, а также в любом отделении связи по каталогу «Роспечать»: № 48060 — годовой индекс, № 15385 — полугодовой индекс.

ЛР № 010292 от 18.08.98.
Сдано в набор 07.05.13. Подписано в печать 10.06.13. Формат 60×84 1/8.
Бумага офсетная. Гарнитура SchoolBookC. Печать офсетная.
Усл. печ. л. 12,1. Уч.-изд. л. 15,2. Тираж 1000 экз. Заказ 175.

Оригинал-макет изготовлен в редакционно-издательском центре ГУАП.
190000, Санкт-Петербург, Б. Морская ул., 67.

Отпечатано с готовых диапозитивов в редакционно-издательском центре ГУАП.
190000, Санкт-Петербург, Б. Морская ул., 67.

УДК 004.932

МЕТОД ФОРМИРОВАНИЯ 3D-ИЗОБРАЖЕНИЯ СЦЕНЫ ПО ОДНОЙ ФОТОГРАФИИ

Н. Н. Красильников,
доктор техн. наук, профессор
Санкт-Петербургский государственный университет аэрокосмического приборостроения

Описан метод формирования 3D-изображения сцены, основанный на аппроксимации ее центральной проекции (2D-изображения) набором плоскостей с последующей декомпозицией сцены на эти плоскости. Для каждой из плоскостей находится аксонометрическая проекция путем использования имеющейся априорной информации об изображенной сцене и карта глубины. Завершающим шагом описываемого метода является «сборка» 3D-изображения сцены объединением аппроксимирующих ее плоскостей и проекция сцены на экран при заданных условиях наблюдения.

Ключевые слова *— 3D-изображение, 3D-сканирование, карта глубины.*

Введение

При наблюдении трехмерных объектов и сцен, несмотря на то, что в их проекциях на сетчатки глаз координата глубины оказывается утраченной, мы, тем не менее, воспринимаем их объемными. Объясняется это тем, что в этих проекциях, которые представляют собой 2D-изображения, содержится информация, используя которую совместно с априорной информацией, имеющейся у зрителя о наблюдаемых объектах, зрительная кора в той или иной мере восстанавливает утраченную при проецировании информацию о координате глубины. Примерами информации, используемой зрительной корой для определения (оценки) утраченной координаты глубины, являются:

— величина искажений геометрических размеров наблюдаемых объектов, возникающих в результате их центральной проекции на сетчатки глаз;

— перекрытие объектами, близко расположенными к зрителю, объектов, которые расположены от зрителя на больших расстояниях;

— распределение полутеней на криволинейных поверхностях объектов, обусловленное эффектом диффузного отражения света;

— расположение объектов относительно линии горизонта;

— наличие атмосферной дымки и пр.

Эти свойства зрения изучались и использовались в архитектуре и живописи, начиная уже с античных времен. Колонны храма Парфенона, форма которых намеренно выбрана отличной от цилиндрической, являются примером одного из первых практических применений законов перспективы в архитектуре. Интенсивное развитие живописи в эпоху Возрождения также в значительной степени связано с использованием перечисленных выше свойств зрения.

В настоящее время возрос интерес к 3D-технологиям и, в частности, к методам получения трехмерных изображений, что, естественно, проявилось в появлении множества публикаций на эту тему, например [1–7], а затем и в разработке 3D-сканеров, 3D-кинотеатров и 3D-телевидения.

В последних двух случаях речь идет по существу о стереоскопических системах. В связи с тем, что изначально съемка 3D-фильмов является весьма дорогостоящим предприятием, появились попытки разработать методы, позволяющие преобразовывать обычные 2D-контенты в 3D-контенты, основанные на использовании перечисленных выше свойств зрения. Так, в университете Карнеги — Меллона был разработан метод преобразования 2D-изображений в 3D-сцену, подробности которого не разглашаются, но известно, что 3D-изображение строится с использованием только вертикальных и горизонтальных поверхностей, обнаруженных на 2D-изображении [8]. Были также разработаны методы, примененные в настоящее время в телевизорах, в которых использованы свойства взаимного перекрытия объектов переднего и последующего планов, а также разница в цвете и контрасте

между изображениями объектов различных планов. Однако качество получаемых при этом изображений невелико. В данной статье рассматривается метод формирования 3D-изображения сцены и составляющих ее объектов, которые можно аппроксимировать плоскостями, по одной 2D-фотографии.

Перспективные преобразования прямых линий

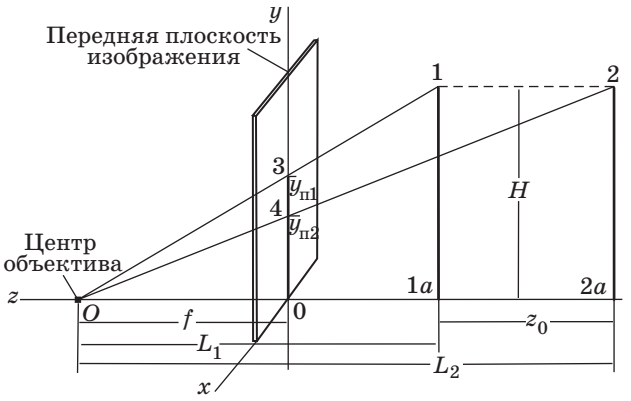
Приведем свойства перспективных преобразований для прямых линий.

1. В общем случае прямая линия на объекте проецируется в прямую линию на изображении центральной проекции, но с *неравномерной шкалой*. Например, движение точки с постоянной скоростью по прямой линии на объекте отображается ее движением на изображении тоже по прямой линии, но с неравномерной скоростью. Чем больше точка удаляется, тем меньше становится видимая скорость ее движения. Эта неравномерность шкалы приводит к тому, что при перспективном преобразовании не сохраняется соотношение длин и площадей в пространстве и в центральной проекции.

2. Параллельные прямые линии в пространстве изображаются в центральной проекции прямыми линиями, сходящимися в одной точке, так называемой точке схода. Если эти линии параллельны горизонтальной плоскости, то их точка схода располагается на линии горизонта. Если они параллельны наклонной восходящей плоскости, то точка их схода располагается выше линии горизонта. Если же они параллельны наклонной нисходящей плоскости, то их точка схода располагается ниже линии горизонта.

Исключением из этого правила является случай, когда прямые линии в пространстве параллельны плоскости, на которую они проецируются. При этом на изображении эти линии также остаются параллельными и не имеют точки схода, а шкала проекции остается равномерной.

Продолжая рассмотрение, выберем правостороннюю систему координат и обратимся к рис. 1, на котором показана так называемая модель камеры с передней плоскостью изображения. Рассмотрим центральную проекцию двух простейших объектов, изображенных на этом рисунке (линий *1-1a* и *2-2a*), на плоскость, при этом координатную ось z декартовой системы координат совместим с осью камеры, а оси x и y — с плоскостью изображения. На рисунке использованы следующие обозначения: H — длина линий, одинаковая для обеих линий; L_1 — расстояние между линией *1-1a* и центром объектива; L_2 — расстояние между линией *2-2a* и центром объектива;



■ ***Рис. 1.*** Модель камеры с передней плоскостью изображения

f — фокусное расстояние объектива; z_0 — расстояние между линиями *1-1a* и *2-2a*; $y_{п1}$ — длина проекции линии *1-1a* на переднюю плоскость изображения (плоскость проецирования); $y_{п2}$ — длина проекции линии *2-2a*.

Из подобия треугольников *1-1a-O* и *3-0-O* следует, что

$$y_{п1}/f = H/L_1, \qquad (1)$$

а из подобия треугольников *2-2a-O* и *4-0-O* следует, что

$$y_{п2}/f = H/L_2. \qquad (2)$$

В общем случае длина проекции $y_п$ линии протяженностью H, которая параллельна плоскости проецирования независимо от ее углового положения по отношению к координатным осям x, y и удалена от центра объектива на расстояние L, равна

$$y_п = H f/L, \qquad (3)$$

где f/L можно рассматривать как масштабирующий множитель. Отсюда следует, что любая точка объекта с координатами $x_{об}$, $y_{об}$ отобразится на центральной проекции точкой с координатами $x_п$, $y_п$:

$$x_п = x_{об} f/L; \;\; y_п = y_{об} f/L. \qquad (4)$$

Из формулы (3), в частности, следует, что при $L = f$ длина проекции линии равна длине самой проецируемой линии. Решая совместно уравнения (1) и (2), найдем расстояние по глубине между двумя линиями *1-1a* и *2-2a*

$$z_0 = L_1 \frac{y_{п1} - y_{п2}}{y_{п2}}. \qquad (5)$$

Рассмотрев основные аспекты перспективного преобразования прямых линий, можно перейти к перспективным преобразованиям плоского прямоугольника.

Перспективные преобразования плоского прямоугольника

Рассмотрим теперь задачу определения истинных размеров одного из простейших объектов, которым является плоский прямоугольник, по его центральной проекции. Будем считать, что две его стороны вертикальны и, следовательно, параллельны плоскости проекции, которая перпендикулярна горизонтальной плоскости, а плоскость, в которой он лежит, повернута на некоторый угол α_0 относительно направления главного луча. В этом случае проекция прямоугольника будет иметь вид трапеции, поскольку размер проекции удаленной стороны вследствие перспективных искажений будет меньше размера ближайшей стороны. Примем также, что нам известны фокусное расстояние объектива f, использованного при фотографировании, расстояние L_1, с которого производилось фотографирование, а также размеры проекций ближайшей и удаленной сторон, соответственно $y_{п1}$ и $y_{п2}$, которые определяются по изображению.

Переходя к определению размеров прямоугольника, найдем размер его вертикальных сторон: $H = y_{п1}L_1/f$, что следует из формулы (1). Для определения размера его горизонтальных сторон, который обозначим $L_{об}$, найдем вначале расстояние по глубине z_0 между этими сторонами по формуле (5). Поскольку плоскость прямоугольника составляет с главным лучом угол α_0, то глубина представляет собой размер проекции его ширины на вертикальную плоскость, в которой лежит главный луч. Отсюда следует, что

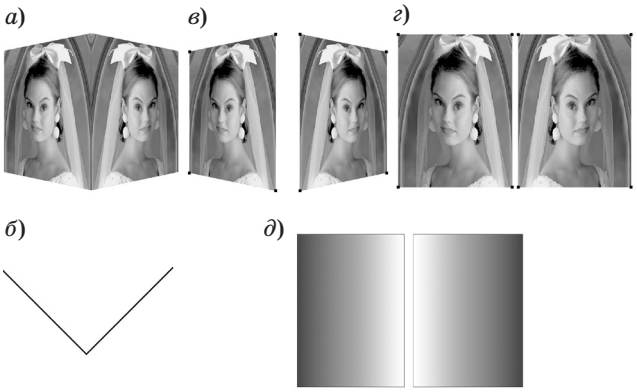
$$L_{об} = z_0/\cos\alpha_0 . \qquad (6)$$

Опуская вследствие громоздкости вычислений рассмотрение случая, когда проецируемой фигурой является плоский многоугольник, стороны которого не обязательно параллельны плоскости, на которую он проецируется, отметим лишь, что методика анализа и в этом случае полностью сохраняется.

Рассмотренный случай, несмотря на свою простоту, важен при решении ряда практических задач, поскольку на практике часто встречаются объекты, оболочки которых могут быть аппроксимированы плоскостями прямоугольной формы. К таким объектам относятся наружные стены зданий, стены интерьеров, предметы мебели, а также ряд других объектов.

Метод формирования 3D-изображения объекта по его двумерной фотографии

Пояснение метода проведем на примере несложного объекта, представляющего собой два



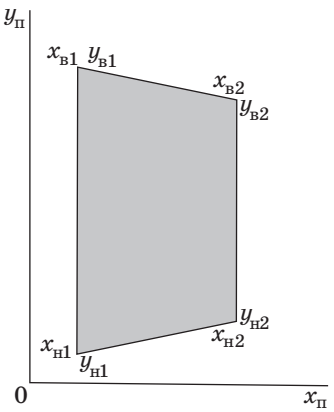
■ ***Рис. 2.*** К пояснению принципа формирования 3D-изображения объекта по его двумерной фотографии: *а* — вид на объект спереди (его фотография); *б* — вид сверху; *в* — две аппроксимирующие плоскости; *г* — аксонометрические проекции каждой из аппроксимирующих плоскостей; *д* — карты глубины каждой из двух аппроксимирующих плоскостей

соединенных прямоугольных рекламных щита (рис. 2, *а*, *б*). Будем считать, что вертикальные линии прямоугольных щитов параллельны плоскости проецирования, а расстояние между ближайшей точкой объекта, т. е. местом, где щиты соединяются, и центром объектива с фокусным расстоянием f равно L_1.

Суть метода заключается в следующем.

1. Вначале выполняется декомпозиция изображения объекта на две аппроксимирующие плоскости (рис. 2, *в*).

2. Далее для каждой из аппроксимирующих плоскостей производится пересчет от перспективной проекции к аксонометрической проекции. Для пояснения того, как это делается, обратимся к рис. 3, на котором показана центральная проекция правой плоскости щита, изображенного на рис. 2, *в*. Из рис. 3 видно, что размер центральной проекции дальней вертикальной гра-



■ ***Рис. 3.*** Центральная проекция прямоугольной плоскости, линии вертикальных границ которой параллельны плоскости изображения

ницы прямоугольного щита $y_{в2} - y_{н2}$ меньше размера центральной проекции его ближней вертикальной границы $y_{в1} - y_{н1}$. Величина этого изменения составляет $K_{min} = (y_{в2} - y_{н2})/(y_{в1} - y_{н1})$.

В такой же мере имеет место уменьшение и горизонтального размера центральной проекции прямоугольного щита, хотя это сразу и не бросается в глаза. Из рис. 3 также следует, что коэффициент изменения координат K линейно изменяется с изменением координаты проекции $x_п$, поскольку уменьшению $x_п$ соответствует уменьшение расстояния между рассматриваемой точкой и центром объектива, т. е.

$$K = K_{min} - (K_{min} - 1)\frac{x_{в2} - x_п}{x_{в2} - x_{в1}}. \qquad (7)$$

Поэтому при определении координат аксонометрической проекции каждой точки объекта на переднюю плоскость изображения $x_а$ и $y_а$ необходимо это учесть следующим образом:

$$y_0 = (y_{в1} + y_{н1})/2;\ \ x_а = x_{в1} + (x_п - x_{в1})/K;$$
$$y_а = y_0 + (y_п - y_0)/K. \qquad (8)$$

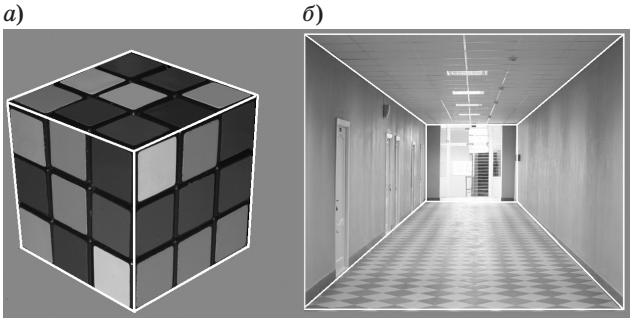
В результате этих расчетов находится аксонометрическая проекция объекта в масштабе, который определяется масштабирующим множителем f/L_1. Координаты $x_{об0}$ и $y_{об0}$ произвольных точек объекта находятся путем деления координат $x_а$ и $y_а$ на масштабирующий множитель f/L_1. Полученные таким образом аксонометрические проекции каждой из аппроксимирующих плоскостей представлены на рис. 2, *г*.

3. Затем определяется расстояние по глубине z_0 между вертикальными линиями, ограничивающими прямоугольные аппроксимирующие плоскости, по формуле (5). Другими словами, определяется расстояние по глубине между двумя плоскостями, параллельными передней плоскости изображения, в которых лежат эти линии.

4. После этого необходимо определить координату глубины для каждой точки аппроксимирующей плоскости, т. е. рассчитать карту глубины. С этой целью вначале методом линейной интерполяции находятся координаты глубины вдоль прямых линий, соединяющих вершины аппроксимирующих плоскостей, для которых координаты глубины известны (в рассматриваемом примере, поскольку плоскости вертикальны, для всех точек вертикальных линий координата глубины будет неизменной). Линейная интерполяция оказывается возможной, поскольку в аксонометрической проекции шкала «дальности», в отличие от шкалы «дальности» центральной проекции, равномерная.

5. Далее методом линейной интерполяции вычисляются координаты глубины для всех точек (пикселей) аппроксимирующей плоскости и таким образом формируются карты глубины каждой из двух аппроксимирующих плоскостей (рис. 2, *д*). На картах глубины более светлый тон соответствует точкам, находящимся ближе к центру объектива, а более темный — точкам, расположенным дальше.

а) *б*)



■ ***Рис. 4.*** Изображения объектов, для аппроксимации которых необходимо использовать 3 (*а*) и 5 (*б*) плоскостей

6. И, наконец, завершающим шагом описываемого метода является «сборка» объекта путем объединения аппроксимирующих его плоскостей.

В зависимости от сложности и количества объектов, изображенных на фотографии, может потребоваться большее число аппроксимирующих плоскостей, чем в рассмотренном примере (рис. 4, *а*, *б*).

Найденные аксонометрическая проекция объекта и карта глубины содержат всю информацию о *видимой части* 3D-объекта и позволяют выводить его изображение на экран дисплея для рассматривания при различных ракурсах и с различных расстояний. Для получения полного 3D-изображения объекта, а не только видимой его части, как и в случае традиционных 3D-сканеров, возникает необходимость в склейках частей 3D-изображений, полученных при различных ракурсах съемки. Обусловлено это тем, что ни один непрозрачный физический объект невозможно одновременно видеть со всех сторон.

Представление 3D-изображения в виде аксонометрической проекции и карты глубины, так называемый 2D+Z-формат, применяют в 3D-телевидении (точнее, в стереоскопическом телевидении). Этот формат может быть преобразован в другие форматы 3D-изображений.

Отображение 3D-изображения на экране дисплея

Покажем теперь, как, располагая аксонометрической проекцией объекта и картой глубины, вывести на экран дисплея центральную проекцию объекта при заданном ракурсе и расстоянии наблюдения. Задача заключается в определении

координат каждой точки формируемого изображения 3D-объекта, соответствующего новому ракурсу наблюдения этого объекта и новому расстоянию, с которого он наблюдается. Будем считать известными расстояние, с которого была произведена съемка L_1, и фокусное расстояние объектива f. Решая задачу, вначале находим координаты объекта, затем осуществляем поворот объекта на заданные углы и, наконец, находим центральную проекцию объекта, видимого с заданного расстояния наблюдения.

Определение декартовых координат x и y точек объекта по его аксонометрической проекции. Как уже было отмечено, декартовы координаты произвольной точки объекта $x_{об0}$, $y_{об0}$, соответствующие его исходному положению, находятся путем деления координат $x_а$ и $y_а$ на масштабирующий множитель f/L_1, т. е.

$$x_{об0} = x_а L_1/f;\ \ y_{об0} = y_а L_1/f. \qquad (9)$$

Что касается координаты глубины произвольной точки объекта $z_{об0}$, то она полностью определяется его картой глубины.

Определение декартовых координат точек объекта, изменившихся в результате его поворота вокруг координатных осей x и y на заданные углы. Поворот объекта вокруг оси z в рассматриваемом случае интереса не представляет, так как его вращение в плоскости проекции не приводит к изменению ракурса наблюдения. Возьмем точку на поверхности объекта с координатами $x_{об0}$, $y_{об0}$, $z_{об0}$, повернем объект сначала на угол α вокруг оси y и найдем его новые координаты $x_{об1}$, $y_{об1}$, $z_{об1}$:

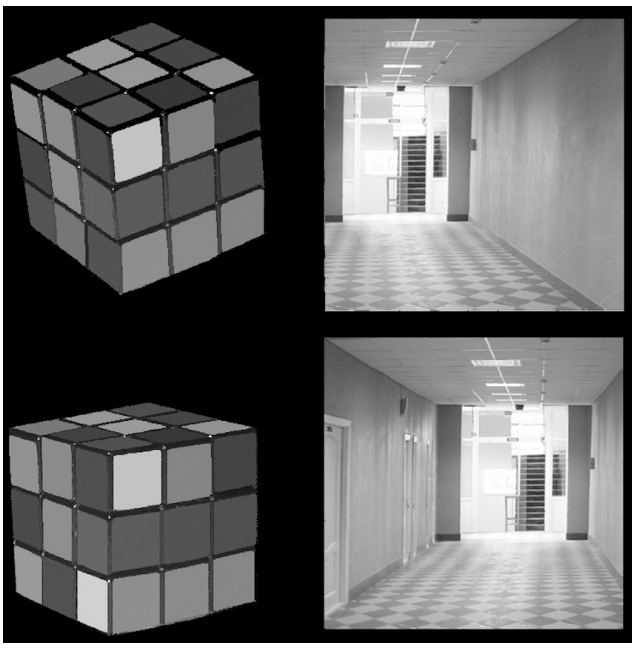
$$x_{об1} = x_{об0}\cos\alpha - z_{об0}\sin\alpha;\ \ y_{об1} = y_{об0};$$
$$z_{об1} = z_{об0}\cos\alpha + x_{об0}\sin\alpha. \qquad (10)$$

Допустимое значение угла α ограничивается значениями, при которых изображение повернутого 3D-объекта еще не включает в себя тех фрагментов, которые при его фотографировании были заслонены самим объектом. Далее, возьмем эту же точку на поверхности объекта после его поворота по азимуту на угол α вокруг оси y и найдем ее новые координаты $x_{об2}$, $y_{об2}$, $z_{об2}$ после поворота объекта вокруг оси x на угол места β:

$$x_{об2} = x_{об1};\ \ y_{об2} = y_{об1}\cos\beta - z_{об1}\sin\beta;$$
$$z_{об2} = y_{об1}\sin\beta + z_{об1}\cos\beta. \qquad (11)$$

Подставив значения $x_{об1}$, $y_{об1}$, $z_{об1}$ из формулы (10) в формулу (11), получим

$$x_{об2} = x_{об0}\cos\alpha - z_{об0}\sin\alpha;$$
$$y_{об2} = y_{об0}\cos\beta - z_{об0}\cos\alpha\sin\beta - x_{об0}\sin\alpha\sin\beta; \qquad (12)$$
$$z_{об2} = y_{об0}\sin\beta + z_{об0}\cos\alpha\cos\beta + x_{об0}\sin\alpha\cos\beta.$$



■ ***Рис. 5.*** Примеры проекций 3D-изображений, полученных описанным методом

Переходя к заключительной операции вычислений, найдем декартовы координаты $x_{пр}$ и $y_{пр}$ центральной проекции объекта на экран дисплея. Согласно формуле (4), между искомыми координатами центральной проекции объекта и соответствующими координатами объекта имеют место соотношения

$$x_{пр} = x_{об2} f/L;\ \ y_{пр} = y_{об2} f/L, \qquad (13)$$

где L — расстояние от центра объектива до точки объекта, координаты которой рассчитываются. Это расстояние равно сумме расстояний: расстояния $L_н$ между центром объектива и ближайшей к камере точкой объекта и расстояния $z_{гл}$ от ближайшей точки к центру объектива до рассматриваемой точки, которое определяется из карты глубины, т. е. $L = L_н + z_{гл}$.

С учетом изложенного имеем

$$x_{пр} = x_{об2}\frac{f}{L_н + z_{гл}};\ \ y_{пр} = y_{об2}\frac{f}{L_н + z_{гл}}. \qquad (14)$$

Проекции 3D-изображений (см. рис. 4), полученные описанной выше обработкой 2D-изображений, показаны на рис. 5.

Заключение

Описанный метод может найти свое применение в тех случаях, когда требуется, используя цифровую камеру, получать 3D-изображения отдельных объектов и целых сцен, которые допускают аппроксимацию плоскостями. Это может

быть актуальным, если принять во внимание, что стоимость лазерного 3D-сканера, используемого для этой цели, в сотни раз превышает стоимость недорогой цифровой камеры.

Поскольку в результате применения описанного выше метода из 2D-изображений формируются 3D-изображения, то не составляет труда сделать следующий шаг и из 3D-изображений получать их стереопары. Опуская вследствие ограниченности объема статьи рассмотрение метода формирования стереопары из 3D-изображения, приведем в качестве иллюстрации лишь сами полученные в процессе выполнения данной работы стереоскопические изображения. На сайте [9] показаны анаглифные изображения, сформированные в конечном итоге из 2D-изображений, которые приведены в тексте настоящей статьи. Анаглифный метод представления стереоскопических изображений для данной демонстрации выбран по причине его доступности. Напомним, что для рассматривания такого рода стереоскопических изображений требуются анаглифные очки, которые в настоящее время чрезвычайно широко распространены и недороги.

Описанный в статье метод может быть также использован как элемент алгоритма, предназначенного для реализации функции преобразования 2D-изображений в стереоскопические изображения, которую начинают применять в современных телевизорах [10].

Литература

1. **Horry Y., Anjyo K., Arai K.** Tour into the picture: using a spidery mesh interface to make animation from a single image // ACM SIGGRAPH Proc. 1997. P. 225–232.
2. **Shum H.-Y., Han M., Szeliski R.** Interactive construction of 3D models from panoramic mosaics // IEEE Conf. on CVPR, June 1998. P. 427–433.
3. **Horn B. K. P.** Height and gradient from shading // Int'l J. of Computer Vision. 1990. Vol. 5. N 1. P. 37–75.
4. **Красильников Н. Н.** Метод получения 3D-изображений, основанный на диффузном отражении света сканируемыми объектами // Информационно-управляющие системы. 2009. № 6(43). С. 7–11.
5. **Красильников Н. Н., Красильникова О. И.** Получение трехмерного изображения объекта путем измерения интенсивности диффузного отражения света различными точками его поверхности // Оптический журнал. 2010. Т. 77. № 6. С. 19–24.
6. **Красильников Н. Н., Красильникова О. И.** Исследование погрешностей определения координаты глубины при 3D-сканировании методом, основанном на диффузном отражении света // Информационно-управляющие системы. 2012. № 3(58). С. 2–8.
7. **Красильников Н. Н.** Цифровая обработка 2D- и 3D-изображений: учеб. пособие для вузов. — СПб.: БХВ-Петербург, 2011. — 608 с.
8. **http:**//www.techeblog.com/index.php/tech-gadget/feature-university-researchers-develop-method-to-convert-2d-images-into-3d-scenes-video- (дата обращения: 02.04.2013).
9. **http:**//www.i-us.ru/authors/krasilnikov_nn.
10. **http:**//3dtv-obzor.ru/2d_3d_convertacia (дата обращения: 02.04.2013).

УДК 621.391

ВОССТАНОВЛЕНИЕ ДВУМЕРНЫХ ИЗОБРАЖЕНИЙ С ДЕФЕКТАМИ

В. Ю. Зиняков,
аспирант
Санкт-Петербургский государственный политехнический университет
А. Е. Городецкий,
доктор техн. наук, профессор
А. Ю. Кучмин,
канд. техн. наук
Институт проблем машиноведения РАН, г. Санкт-Петербург
Е. И. Зеленев,
доктор ист. наук, профессор
Н. В. Алферова,
доцент
Санкт-Петербургский государственный университет

Рассматривается задача обработки и восстановления изображений типа графический орнамент. Предложен новый подход, который заключается в комбинировании известных математических алгоритмов для достижения бóльших надежности и вычислительной экономичности алгоритма восстановления. Метод проходил тестирование и был оптимизирован для анализа и последующего восстановления исторических орнаментов, составленных главным образом из геометрических мотивов.

Ключевые слова *— обработка изображений, вейвлет-анализ, фрактальный анализ.*

Введение

Проблема восстановления изображений с дефектами актуальна при исследовании поврежденных исторических орнаментов на основе статистических данных и информации, полученной из уцелевших фрагментов. Разработано большое количество методов распознавания и восстановления изображений, основанных на учете их семантической структуры [1, 2]. К сожалению, подобные методы не адаптированы к восстановлению и классификации исторических орнаментов. Традиционно подобная задача решается специалистом в данной области, однако до сих пор не было разработано эффективного алгоритма для автоматизации решения данной проблемы. Поиску, разработке и внедрению такого алгоритма и посвящена данная работа.

Объект исследования. Категории орнамента: ритм, стиль

Дословный перевод с латинского слова орнамент — украшение. В специальной литературе орнамент означает узор, состоящий из ритмически упорядоченных элементов для украшения каких-либо предметов или архитектурных сооружений.

Исторически сложилось так, что человек заметил пользу в работе упорядочивающего начала, позволяющего тратить меньше сил, делать нарядным, праздничным быт, в результате чего появилось понятие ритм. Люди стали передавать свои зрительные впечатления в виде изображений — узоров, орнаментов. Для создания элементов орнамента человек выбирал силуэты птиц, рыб, животных, самого себя, растений. Но ритмическая основа орнамента всегда оставалась, а содержание менялось вместе с условиями жизни. Кроме натурных изображений стали использоваться украшения неизобразительного характера, способные вызывать у человека радость, печаль, создавать ощущение покоя.

Позже появилось понимание стиля в смысле единого художественного оформления здания, вещей, предметов, окружающих человека. Стиль как образная система основан на единстве идейного содержания, порождающего единство всех элементов

художественной формы, всех художественно-выразительных средств. В буквальном смысле слово стиль обозначает то видимое, ощутимое своеобразие, которое, прежде всего, бросается в глаза и является отличительным признаком художественного произведения. Понятие это бесконечно многообразно. Различают стиль одного произведения, целого ансамбля, индивидуальный, авторский стиль. Можно говорить о стиле отдельных стран, народов, художественных направлений, например, русский, китайский, строгий, суровый, исторический, романский, готический, модерн, ретро и др.

Элемент орнамента, многократно повторяющийся, называют раппортом. Раппорт — повторяющаяся часть узора.

Мотив — простейшая динамическая смысловая символическая единица орнамента. Кроме того, это слово имеет другое значение — материал для создания сюжета раппорта.

Источником для создания орнамента служат реальный мир, природа, мифология, народный эпос или же используются строгие геометрические фигуры, геометрические построения, письмена, шрифт. В ходе развития орнаментального искусства четко выделились типы орнаментов. Лента, или ленточный орнамент, пожалуй, наиболее многочисленный подвид орнаментов. Он применяется при оформлении зданий, предметов, вещей в качестве бордюра, фриза, каймы, тесьмы, обрамления и др. Один из самых древних, распространенных и богатых вариантами видов ленточного орнамента — меандр.

В данной работе наиболее тщательно исследуются такие раппорты орнаментов, как крест и лента (линия).

Алгоритм восстановления

Для технической реализации решения предложенной задачи в данной работе был использован пакет MatLab [3–5].

Выбор этой среды разработки был обусловлен следующими факторами:

1) оптимизация языка MatLab для технических вычислений и обработки изображений;

2) наличие большого количества готовых программных решений в языке MatLab;

3) наличие обширной документации по языку MatLab: как предложенной разработчиками, так и сторонней;

4) высокая производительность (при должной оптимизации) программы на языке MatLab.

Алгоритм регенерации орнаментов включает в себя следующие шаги.

1. Считывание данных из файла и команд пользователя посредством графического интерфейса пользователя.

2. Удаление областей повреждения.

3. Конвертирование исходного изображения в бинарное.

4. Разбиение изображения на мотивы.

5. Нахождение дочерне-родительских связей между мотивами.

6. Нахождение одинаковых мотивов.

7. Нахождение углов поворота одинаковых мотивов относительно друг друга.

8. Нахождение композиционных фигур на изображении (композиционный анализ).

9. Регенерация композиционных фигур.

10. Синтез композиционных фигур на основе имеющихся.

Шаги 3–6 при этом условно названы графическим анализом.

Удаление областей повреждения

Алгоритм не производит анализа указанных пользователем поврежденных областей. Данные об их границах хранятся в программе для последующего использования, а поврежденная область на данном этапе заполняется белым (если изображение требуется инвертировать) либо черным (в противном случае) фоном.

Конвертирование исходного изображения в бинарное

Графический и композиционный анализ исходного изображения целесообразно производить на бинарном изображении, без учета его цветности и насыщенности. Для приведения к бинарному виду над изображением выполняются следующие действия.

1. Приведение исходного цветного (*true-color*) изображения к формату «оттенки серого» (*gray-scale*) путем игнорирования оттенка (*hue*) и насыщенности (*saturation*) пиксела с сохранением его яркости (*luminance*).

2. Нахождение порогового значения границы черное—белое по методу Оцу (по умолчанию) либо получение желаемой пользователем границы посредством интерфейса.

3. Приведение изображения в формате «оттенки серого» к бинарному формату при использовании найденного порогового значения.

4. Инвертирование изображения.

Обоснование инвертирования изображения

Здесь и далее в работе белые (единичные) фрагменты бинарного изображения будут называться узором орнамента, а черные (нулевые) фрагменты — его фоном.

Как графический, так и композиционный анализ и последующий синтез изображения осуществляются при наличии черного фона (0 в бинарной матрице) и белого узора орнамента (1 в би-

нарной матрице). Поскольку статистика по орнаментам показывает, что в подавляющем большинстве случаев анализировать требуется орнамент, построенный при помощи темного узора на светлом фоне, полученное бинарное изображение требуется инвертировать. Это является настройкой по умолчанию, которую пользователь может изменить.

Разбиение изображения на мотивы

Мотивом называется фрагмент изображения, который изолирован либо имеет незначительное количество точек соприкосновения с прочим узором орнамента. Алгоритм восстановления орнаментов, рассматриваемый в данной статье, построен на следующих важных свойствах мотива:

1) мотив визуально изолируется наблюдателем-человеком;

2) мотив может состоять из других мотивов, по отношению к нему называемых дочерними;

3) в изображении может содержаться множество одинаковых мотивов. Мотив A считается одинаковым по отношению к мотиву B, если его можно получить из мотива B путем применения к последнему таких аффинных преобразований, как поворот и зеркальное отражение;

4) как правило, орнамент состоит из одинаковых мотивов, объединенных в композиционные фигуры.

Для разбиения орнамента на мотивы используется следующий итеративный алгоритм.

1. Если это не первая итерация алгоритма, над изображением производится шаг приближения к узловым (опорным) точкам скелета изображения.

2. Над изображением проводится операция *H*-разбиения.

3. Находятся полностью изолированные фрагменты узора (мотивы).

4. Если это не первая итерация алгоритма, найденные мотивы восстанавливаются до первоначальных значений.

5. Мотивы, площадь которых меньше заданной константы, считаются незначительными (возможно, артефактами изображения) и не принимаются в расчет при дальнейших исчислениях. Площадью мотива здесь и далее называется количество его единичных элементов.

6. Если данный мотив не был учтен на предыдущей итерации, добавляем его в список найденных мотивов.

7. Инкремент счетчика итераций и возврат к шагу 1.

Количество итераций данного алгоритма задается пользователем (по умолчанию равно трем) и определяет степень (глубину) разбиения мотивов на более мелкие.

Обращаем также ваше внимание на то, что некоторые найденные мотивы являются дочерними мотивами других.

Приближение к узловым точкам скелета

Скелетом бинарного изображения называется изображение, полученное из исходного путем итеративного обнуления границ узора, но таким образом, чтобы ранее цельные объекты оставались цельными. Ширина итоговых объектов составляет 1 пиксель.

Узловыми точками скелета называются единичные точки, находящиеся на пересечении линий скелета. Например, показанному на следующей матрице скелету соответствует следующая узловая точка:

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Процесс приведения изображения к его узловым точкам итеративный. В настоящей работе используется только несколько (как правило, 3) его итераций, для того чтобы разделить орнамент на изначально соединенные между собой мотивы.

Восстановление мотивов

Поскольку в результате разделения мотивов они искажены относительно оригинальных, так как их площадь уменьшена, требуется восстановить мотивы до оригинальных размеров. Следовательно, те граничные (контурные) точки узора, которые были обнулены, должны быть отнесены к тому или иному мотиву. Для этого выполняются следующие действия.

1. Создается нулевая бинарная матрица тех же размеров, что и исходное изображение.

2. На нее помещаются найденные искаженные мотивы, причем элементы каждого из помещенных мотивов равны порядковому номеру мотива, начиная с 1. На этом этапе матрица состоит из нулей (фон) и значений от 1 до N (N — количество найденных мотивов).

3. Полученная матрица складывается с исходным, неразделенным на мотивы бинарным изображением. На этом этапе матрица состоит из нулей (фон), значений от 2 до $N + 1$ (искаженные мотивы) и единиц (пикселы, которые требуется соотнести с тем или иным мотивом).

4. Для каждого единичного элемента полученной матрицы находится ближайший элемент, больший 1. Алгоритм нахождения такого пиксела заключается в поиске такого элемента вокруг

исходной точки, с преобразованием и последующим округлением полярных координат и постепенным увеличением радиуса поиска.

Алгоритм порождает погрешности ввиду округления и выбора большого шага для угла φ. Исследования показали, что погрешности на данном этапе вполне допустимы, поскольку задача соотнесения пикселов к мотивам (и разделение мотивов как таковое) сложно выражаема естественным языком и допускает вариации при формализации. Неоспоримым достоинством метода является его низкая временная ресурсоемкость.

5. Все ненулевые элементы матрицы декрементируются. Полученная матрица идентична исходной по местонахождению узора и фона (за исключением найденных артефактов), но содержит разные значения для разных мотивов (т. е. на данном этапе они изолированы).

Нахождение дочерне-родительских связей между мотивами

Мотив A называется дочерним по отношению к мотиву B, если при покоординатном помещении их обоих на изображение мотив A полностью содержится в мотиве B.

На практике используется следующее условие: если 95 % площади мотива A содержатся на площади мотива B, мотив A является дочерним по отношению к B. Если мотив A является дочерним по отношению к B, а B — к C, то A является дочерним по отношению к C (транзитивность).

Вероятностные коэффициенты здесь и далее были подобраны эмпирическим путем, и в последующих шагах алгоритма восстановления изображений решения будут применяться с возрастающей степенью допущения (т. е. с меньшей точностью), однако, как будет показано впоследствии, благодаря историческим статистическим данным это не повлияет на точность восстановления изображения.

Нахождение одинаковых мотивов

Мотивы A и B называются одинаковыми, если мотив A может быть получен из мотива B путем таких аффинных преобразований, как поворот и зеркальное отражение. Мотивы, имеющие ту же форму, но разный размер, в данном методе не считаются одинаковыми. Предполагается, что в орнаменте существует несколько множеств одинаковых элементов. На данном этапе требуется определить эти множества и входящие в них элементы.

Описанные ниже методы являются во многом эмпирическими: как входящие в них константы, так и некоторые условия принятия решения получены опытным путем на реально существующих исторических орнаментах.

Справедливо следующее высказывание, определяющее первичный негативный признак сортировки:

• если мотив A является дочерним по отношению к мотиву B, они не могут считаться одинаковыми.

Простейшим в программной реализации и наиболее точным методом обнаружения одинаковых мотивов, а также углов их поворота относительно друг друга является метод пошагового поворота мотива A вокруг заданной точки (к примеру, центра масс мотива) и вычисления корреляционной функции мотивов A и B на каждом шаге. Метод используется в данной работе, однако его существенным недостатком является очень низкая производительность. Для повышения общей производительности метода в алгоритм включены предварительные действия, предназначенные для уменьшения размерности данных.

В математической статистике различают ошибки I и II рода. Если мотив a принадлежит к множеству A, но отвергнут алгоритмом (неверно отвергнут), то это называют ошибкой I рода. Если мотив a не принадлежит к множеству A, но принят алгоритмом (неверно принят), то это называют ошибкой II рода.

Поскольку вышеуказанные шаги снижают размерность данных для алгоритма поворота и вычисления корреляционной функции, который отличается высокой точностью и низкой производительностью, на этом этапе приоритетно снизить ошибку типа I настолько, насколько это возможно, даже допуская при этом рост ошибки II типа.

Это обусловливает выбор высоких «коэффициентов доверия» в описанных ниже методах предварительной сортировки.

Сортировка по площади

Ниже описан негативный признак принадлежности мотивов к одному множеству. Справедливо следующее утверждение (напомним, площадью бинарного мотива называется количество его единичных пикселов):

• если отношение модуля разности сумм мотивов к среднему арифметическому сумм этих мотивов больше константы, равной 0,2, эти мотивы не являются одинаковыми, т. е.

$$|S_A - S_B| / \frac{S_A + S_B}{2} > 0{,}2,\ A \neq B.$$

Сортировка по главным осям мотива

Для дальнейшей сортировки требуется найти большую и малую оси мотива — отрезки, отвечающие следующим требованиям:

— большая ось мотива — это наибольший возможный отрезок, проходящий через центр масс

мотива и соединяющий две точки внешнего контура мотива;

— малая ось мотива — это отрезок, перпендикулярный большой оси, проходящий через центр масс мотива и соединяющий две точки внешнего контура мотива.

Центром масс мотива называется точка, координаты которой находятся по следующей формуле:

$$x = \mathrm{sum}(x)/N,\ \ y = \mathrm{sum}(y)/N.$$

Дальше будет проведена отдельная сортировка по трем различным методам. Результаты сортировки по каждому из этих методов сохраняются отдельно, и итоговое решение о том, являются ли данные мотивы одинаковыми, будет принято по этим результатам.

Сортировка по длине мотива

Мотивы A и B являются одинаковыми по длине, если справедливо следующее неравенство:

$$|l_{\mathrm{A}} - l_{\mathrm{B}}| / \frac{l_{\mathrm{A}} + l_{\mathrm{B}}}{2} < 0{,}3,\ \mathrm{A} = \mathrm{B},$$

где l_{A}, l_{B} — соответственно длины больших осей мотивов.

Если мотивы равны по длине, на данном этапе также производится определение угла поворота одного мотива относительно другого. Пусть $a_1(x_{a1}, y_{a1})$, $a_2(x_{a2}, y_{a2})$ — точки пересечения главной оси мотива A и его внешнего контура [4]. Тогда

$$\mathrm{tg}_{\mathrm{A}}\alpha = \frac{y_{a2} - y_{a1}}{x_{a2} - x_{a1}}.$$

Аналогично найдем $\mathrm{tg}_{\mathrm{B}}\beta$ для мотива B. Тогда угол поворота мотива B относительно мотива A

$$\varphi_{\mathrm{B,\,A}} = \mathrm{arctg}(\mathrm{tg}_{\mathrm{B}}\beta) - \mathrm{arctg}(\mathrm{tg}_{\mathrm{A}}\alpha) + \pi k,\ k \in \mathbb{Z}.$$

Сортировка по матрице поворота

В данном методе проверяется возможность получения трех граничных точек мотива B из трех граничных точек мотива A путем преобразования Гивенса.

Граничными точками мотива называются точки пересечения его главной оси с внешним контуром, а также точка пересечения малой оси с внешним контуром (в целях уменьшения погрешности берется та точка, которая более удалена от центра масс). Тогда справедливо равенство

$$\mathbf{AM} = \mathbf{B},\ \ \mathbf{M} = \mathbf{BA}^{-1},$$

$$\mathbf{A} = \begin{pmatrix} x_{a1} & x_{a2} & x_{a3} \\ y_{a1} & y_{a2} & y_{a3} \\ 1 & 1 & 1 \end{pmatrix};\ \mathbf{B} = \begin{pmatrix} x_{b1} & x_{b2} & x_{b3} \\ y_{b1} & y_{b2} & y_{b3} \\ 1 & 1 & 1 \end{pmatrix}.$$

z-координата точек принята за 1, чтобы матрица **A** была обратимой. Тогда, если мотивы A и B являются одинаковыми, матрица **M** имеет вид

$$\mathbf{M} = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

По формулам легко можно проверить, так ли это. Однако, несмотря на кажущийся высокий потенциал метода, погрешность на практике очень велика. В итоге при реализации метода проверяются только следующие факторы:

1) модуль каждого элемента матрицы **M** не должен превышать 2 (если он больше 1, элемент в зависимости от знака приравнивается к 1 или –1);

2) элементы в позициях (1, 1) и (2, 2) должны быть одного знака;

3) элементы в позициях (1, 2) и (2, 1) должны быть разных знаков.

Если эти условия выполнены, мотивы считаются одинаковыми по матрице поворота, а угол поворота одного мотива относительно другого вычисляется как среднее арифметическое арккосинуса среднего арифметического элементов (1, 1) и (2, 2) и арксинуса среднего арифметического элементов (1, 2) и (2, 1). Все углы при этом предварительно переводятся в первую четверть, т. е. на практике найденный угол (с учетом погрешности)

$$\varphi_{\text{действ}} \approx \varphi_{\text{выч}} + \frac{\pi k}{4},\ k \in \mathbb{Z}.$$

Сортировка по крестообразному делению

В данном методе мотивы сравниваются по пропорции деления большой оси центром масс мотива. Метод обладает относительно высокой точностью, однако в нем не предусмотрено нахождение угла поворота одного мотива относительно другого.

Условием равенства мотивов по крестообразному делению является следующее неравенство:

$$|k_{\mathrm{A}} - k_{\mathrm{B}}| / \frac{k_{\mathrm{A}} + k_{\mathrm{B}}}{2} < 0{,}3,\ \mathrm{A} = \mathrm{B},$$

где k_{A}, k_{B} — отношения отрезков, на которые центр масс делит большую ось мотива, вычисленные для мотивов A и B.

Определение множества равенства

Множеством равенства называется множество мотивов орнамента, которые в ходе выполнения аргумента были признаны одинаковыми. Мотив считается принадлежащим к множеству равенства, если он принадлежит к этому множеству по результатам хотя бы двух из трех описанных выше алгоритмов, а именно метода равенства по длине, метода равенства по матрице поворота и метода равенства по крестообразному делению.

■ *Рис. 1.* Множества равенства М

Некоторые полученные в результате фильтрации множества равенства показаны на рис. 1. Ошибка типа II на данном этапе не была аннулирована.

Нахождение углов поворота мотивов

Следующим шагом является нахождение угла, на который повернут каждый мотив относительно первого мотива в данном множестве равенства, который называется эталонным.

Для этого используется следующий алгоритм.

Определим возможный диапазон угла поворота. Если для мотива был найден и угол поворота по матрице, и угол поворота по длине, и эти углы (находясь в I четверти; при необходимости из угла предварительно вычитается $\frac{\pi k}{4}, k \in \mathbb{Z}$) отстоят друг от друга больше чем на 40°, границами возможного интервала являются эти углы. В противном случае интервал равен среднему арифметическому этих углов (или единственному найденному углу) ±20°.

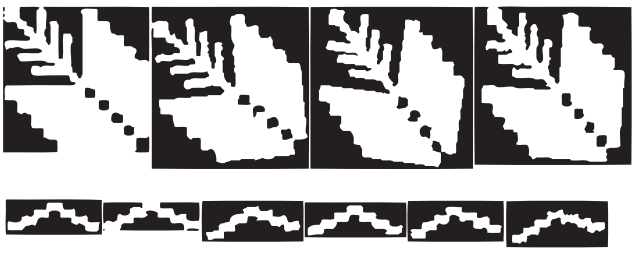
После этого проходим по интервалу с шагом в 5°, поворачивая тестируемый мотив вокруг его центра масс и вычисляя корреляционную функцию этого мотива по отношению к тестовому. Шаг выбран достаточно большим из соображений увеличения производительности метода; как будет показано далее, эта ошибка будет аннулирована.

Корреляционная функция двумерных изображений имеет вид

$$r = \frac{\sum_m \sum_n \left(\mathbf{A}_{mn} - \bar{\mathbf{A}}\right)\left(\mathbf{B}_{mn} - \bar{\mathbf{B}}\right)}{\sqrt{\left(\sum_m \sum_n \left(\mathbf{A}_{mn} - \bar{\mathbf{A}}\right)^2\right)\left(\sum_m \sum_n \left(\mathbf{B}_{mn} - \bar{\mathbf{B}}\right)^2\right)}},$$

и ее значения находятся в пределах $C_2 \in [0; 1]$. После этого интервал проверяется для II четверти, т. е. к верхней и нижней границам интервала прибавляется 90°, затем для III и IV четвертей.

Если на какой-либо итерации корреляционная функция превышает 0,7, алгоритм прерывается и найденный угол считается истинным.



■ *Рис. 2.* Примеры мотивов, повернутых на найденный угол

Если максимальное найденное значение корреляционной функции не превышает 0,3, мотив исключается из множества равенства.

Мотивы, повернутые на найденный угол, показаны на рис. 2.

Композиционный анализ

Следующим шагом является композиционный анализ, т. е. изучение взаимного расположения мотивов на изображении. Исследование орнаментов показало, что наиболее часто встречающимися (в той или иной вариации) композиционными фигурами орнаментов являются линия и крест. Линией в данной работе также является такая статистически частая композиционная фигура, как меандр.

Крестом в данной работе называется такое расположение четырех мотивов, при котором мотивы повернуты на соответственно 90, 180, 270° относительно четвертого, эталонного мотива, а центры масс мотивов удалены от центра масс эталона на, соответственно, $x, \sqrt{2}x, x$ пикселов.

Линией в данной работе называется такое расположение N мотивов, при котором центры масс мотивов лежат на прямой, а расстояние между центрами мотивов равно константе.

Статистический анализ орнаментов показал, что крест — очень частая геометрическая фигура, и если в ходе предыдущих шагов алгоритма были найдены мотивы, даже с большой погрешностью удовлетворяющие описанным выше условиям креста, целесообразно проверить данную гипотезу.

На данном этапе проводится и восстановление неполных крестов. Алгоритм поиска и восстановления следующий.

Для каждого мотива, входящего в какое-либо множество равенства, в том же множестве равенства ищутся мотивы, угол поворота которых относительно этого мотива равен соответственно 90, 180 и 270°. Если найдено хотя бы два таких мотива (из трех), то производится проверка расстояния между центрами масс эталонного мотива и найденных.

Статистический анализ орнаментов показал, что:

1) линии составлены из более мелких мотивов, нежели кресты; это приводит к возрастанию погрешности;

2) составляющие линию элементы не повернуты относительно эталонного.

Ввиду этих факторов в данной работе при поиске линии не учитываются углы поворота мотивов.

Если в ходе композиционного анализа было установлено, куда следует помещать эталонный элемент и какие предварительные преобразования над ним следует проделывать, его размещение не составляет труда. Требуется переместить пикселы цветного изображения (координаты которых совпадают с координатами перемещаемого мотива) на указанную позицию. Единственным предварительным действием является проверка чередования цветности.

Как правило, элементы, входящие в композиционную группу «крест» или «линия», либо имеют ту же цветность, либо их цветовые параметры совпадают для накрест лежащих элементов (для креста) или чередуются через один (для линии).

При проверке цветности отдельно сравнивается среднее значение красного для всего мотива, среднее значение зеленого и синего. Если эти значения для двух разных мотивов отличаются не более чем на 50 (при максимуме в 255), цветность этих мотивов считается одинаковой.

Если чередование цветности для композиционной фигуры установлено, при ее восстановлении может использоваться не эталонный мотив, а смежный с ним (для креста) или ближайший к нему (для линии).

Копирование композиционных фигур

Полное восстановление композиционной фигуры возможно в том случае, если та же фигура, состоящая из тех же мотивов, встречается на орнаменте неоднократно, и одна из них не пострадала (или была полностью восстановлена описанными выше методами), а вторая сохранилась хотя бы частично. Формальное описание условия звучит следующим образом:

• если существует полностью восстановленная композиционная фигура и существует мотив, входящий в то же множество равенства, что и мотивы данной фигуры, но не входящий в фигуру, то целесообразно проверить, не является ли этот мотив частью такой же фигуры, утраченной в результате дефекта.

Проверка данной гипотезы происходит по тому же принципу, что и восстановление композиционной фигуры: по количеству пикселов, которые в результате переноса фигуры попали на фон, а не на узор и поврежденные области.

Данный метод является простым в разработке, нетребовательным к производительности, точным и позволяет восстанавливать значительные части орнамента.

Сложность и оптимизация алгоритма

Значительная часть разработанных и рассмотренных в данной статье алгоритмов имеет сложность $T(n^2)$, поскольку алгоритмы сравнения мотивов по тому или иному признаку содержат два вложенных цикла. Для оптимизации были приняты следующие меры:

1) все возможные метаданные, необходимые для сравнения мотивов, для каждого из мотивов вычисляются однократно и при выполнении сравнения хранятся в оперативной памяти;

2) память своевременно освобождается, когда метаданные становятся не нужны.

Это привело к появлению алгоритмов сложности $T(n)$ с большим количеством флопов на каждой итерации и алгоритмов сравнения сложности $T(n^2)$ с минимально возможным количеством флопов. Алгоритмом сложности $T(n^2)$ со значительным количеством флопов является только алгоритм определения угла поворота; для его оптимизации размерность данных была сокращена.

Хранение дополнительных данных не приводит к снижению производительности, так как разработанный алгоритм обладает высокой временной сложностью, но объемы данных, с которыми он работает, достаточно малы (хотя пространственная сложность также достаточно высока).

Поскольку после обработки единичного орнамента память освобождается, объем оперативного запоминающего устройства не является сдерживающим фактором данного алгоритма. Целью проводимой оптимизации было снижение требуемого процессорного времени, в том числе, как было показано, и за счет снижения точности промежуточных данных, не влияющей на точность результата.

Заключение

Для идентификации и восстановления изображений были использованы следующие известные методы: повышение/снижение яркости, повышение/снижение контрастности, негатив, бинаризация, аффинные преобразования, масштабирование, сглаживающий фильтр, медианный фильтр, дискретное преобразование Фурье, методы выделения границ (метод Робертса, метод Лапласа, метод Уоллеса, метод Собела, метод Кирша, статистический метод), корреляционный и регрессионный анализы. Проведенные исследования показали, что применения классических методов не позволяют с надлежащим качеством решить поставленную задачу. Основным препят-

ствием является наличие паразитных шумов и дефектов изображений, которые могут быть удалены методами обработки изображений, основанных на применении вейвлет и фрактальной фильтрации, совмещаемой с логико-вероятностным анализом результатов.

Эффективность алгоритма во многом зависит от характера орнамента, расположения поврежденных областей, а также от качества предложенного изображения. Для случаев простых геометрических орнаментов, характерных для культур Ближнего и Дальнего Востока (для работы с которыми и был оптимизирован данный алгоритм), площадь поврежденных и впоследствии восстановленных сегментов в среднем должна достигать 40–50 % общей площади изображения.

Литература

1. **Красильников Н. Н.** Принципы обработки изображений, основанные на учете их семантической структуры // Информационно-управляющие системы. 2008. № 1. С. 2–6.
2. **Обухова Н. А.** Предварительная классификация изображения в задачах сегментации объектов // Информационно-управляющие системы. 2007. № 2. С. 22–28.
3. **Смоленцев Н. К.** Основы теории вейвлетов. Вейвлеты в MATLAB. — М.: ДМК Пресс, 2005. — 304 с.
4. **Городецкий А. Е., Тарасова И. Л.** Управление и нейронные сети. — СПб.: СПбГПУ, 2005. — 400 с.
5. **Дьяконов В. П., Абраменкова И. В.** MATLAB. Обработка сигналов и изображений: специальный справочник. — СПб.: Питер, 2002. — 608 с.

УДК 681.3

АНАЛИЗ НАДЕЖНОСТИ ЦИФРОВЫХ УСТРОЙСТВ СО СТРУКТУРНЫМ РЕЗЕРВИРОВАНИЕМ И ПЕРИОДИЧЕСКИМ ВОССТАНОВЛЕНИЕМ РАБОТОСПОСОБНОГО СОСТОЯНИЯ УЗЛОВ

С. Л. Максименко,
старший преподаватель
В. Ф. Мелехин,
доктор техн. наук, профессор
Санкт-Петербургский государственный политехнический университет

Предлагается метод оценки надежности цифровых устройств, представленных в виде сети троированных узлов, учитывающий цикличность вычислительных процессов и периодическое восстановление информации в узлах. Метод основан на разбиении устройства на ячейки с независимыми отказами. Для произвольной ячейки предложена приближенная оценка вероятности безотказной работы для случая, когда период восстановления информации в узлах много меньше среднего интервала между отказами. Получена зависимость интенсивности отказов ячейки от периода восстановления информации и показано, что эта зависимость линейная независимо от структуры связей в ячейке.

Ключевые слова *— цифровые устройства, интегральные схемы, сбои, отказы, периодическое восстановление, надежность, резервирование, оценка, модель, структура, троирование, мажоритар.*

Введение

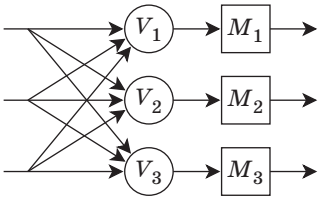
Анализ влияния радиационных эффектов на информационные процессы в цифровых устройствах [1, 2] подтверждает актуальность совместного использования методов повышения устойчивости к радиационным воздействиям на различных уровнях организации системы. Основным методом, применяемым на уровнях функциональной организации и схемотехнической реализации устройств управления и обработки информации, является структурное резервирование узлов с использованием голосующих элементов (мажоритаров) для блокирования отказов.

В статье [2] на основе аналитического обзора работ по влиянию радиации на цифровую аппаратуру установлено, что восстанавливаемые отказы, проявляющиеся в искажении информации, возникают значительно чаще, чем невосстанавливаемые отказы, проявляющиеся в изменении полупроводниковой структуры. Показано также, что существенно повысить надежность устройства можно путем организации циклической работы узлов с периодическим восстановлением информации. В ходе проектирования необходимо осуществить выбор оптимального варианта резервирования с учетом процессов восстановления информации при ограничениях на аппаратные ресурсы и быстродействие устройства.

Для выбора оптимального варианта необходимо иметь механизм построения оценки параметров надежности при заданном варианте резервирования и организации вычислительных процессов.

В работе [3] предложена модель, позволяющая оценить надежность троированного узла при периодическом восстановлении информации. Оценка основана на предположении, что работа узла организована циклически, и с началом нового цикла происходит загрузка в узел новой информации, что обеспечивает восстановление узла после отказа, если он имел место на предыдущем цикле. Показано, что при низкой вероятности отказа за один цикл работы узла и при большом количестве циклов интенсивность потока отказов троированного узла пропорциональна периоду восстановления информации.

Цифровое устройство можно представить как сеть резервированных узлов. Будем рассматри-



■ ***Рис. 1.*** Сеть троированных узлов с троированными мажоритарами

вать широко распространенный на практике вариант резервирования, когда каждый узел троируется и тройки узлов (a_i, c_i, d_i) соединены через троированные мажоритары (b_i). Пример такой системы показан на рис. 1.

Будем считать систему исправной, если в каждой тройке элементов хотя бы два формируют правильное выходное значение. В свою очередь, элемент формирует правильный выход, если он исправен и на его входы поступают правильные значения. Мажоритар формирует правильное значение, если он исправен и хотя бы на два его входа поступают правильные значения.

Оценка надежности сети троированных узлов даже при известных параметрах элементов является сложной задачей, так как отказы узлов в сети в общем случае не являются независимыми.

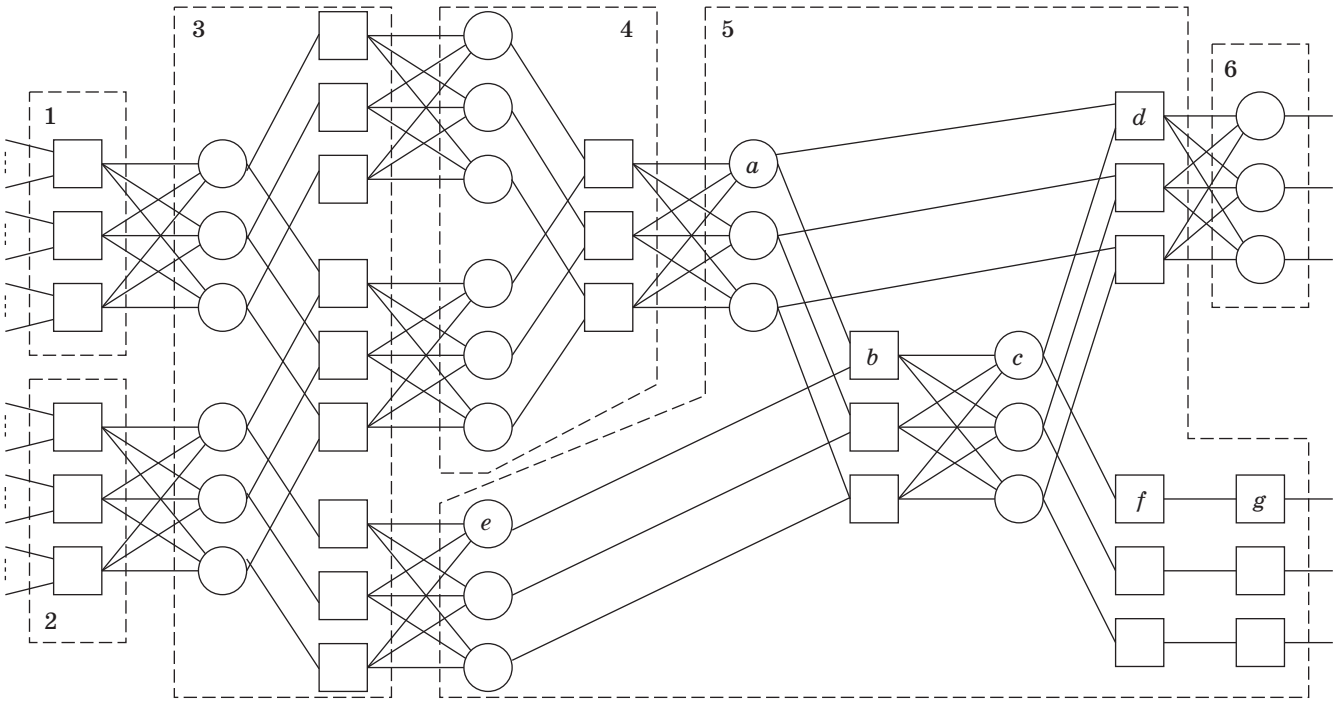
Рассмотрим устройство, структура которого представлена на рис. 1. Сбой в мажоритаре b_1 приводит к отказу узлов c_1 и d_1, причем восстановление информации в узлах произойдет только с началом их следующего цикла работы. Для оценки вероятности исправности системы необходимо определить все сочетания отказов элементов, при которых система остается исправной, и вероятности наступления данных комбинаций отказов. Далее мы построим аналитическую оценку, параметрами которой являются интенсивности отказов элементов и периоды их работы.

Структура сети резервированных элементов

Система, представляющая собой сеть из троированных узлов, соединенных мажоритарами с соблюдением ряда ограничений, может быть разбита на «домены» — «наборы не связанных друг с другом узлов с голосующим устройством на входе и ациклической структурой» [4].

Рассмотрение системы как множества доменов удобно тем, что в силу наличия в доменах голосующих элементов на входе отказы доменов происходят независимо, что позволяет существенно сократить сложность задачи анализа системы: в некоторый момент времени вероятность нахождения системы в исправном состоянии равна произведению вероятностей исправности всех доменов.

Близкий подход, использующий разбиение системы на группы резервированных узлов, предложен в работе [5]. Ее авторы показывают, что любая сеть троированных элементов TMR network может быть разбита на ячейки (TMR cells). На рис. 2 пун-



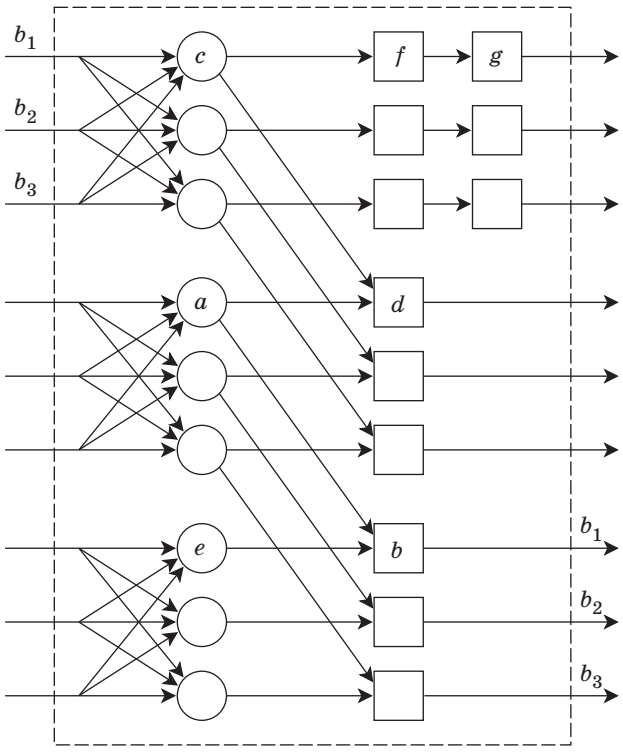
■ ***Рис. 2.*** Сеть, разбитая на ячейки

ктиром приведен пример сети TMR network, где выделены TMR cells *1–6*. Ячейка состоит из троек мажоритаров и троек резервированных модулей. Функция работоспособности ячейки задается как условие, что в каждой тройке модулей, входящей в ячейку, как минимум два сформируют правильное значение (т. е. будут исправны они сами и все мажоритары на их входе). В работе [5] предложен алгоритм численного расчета надежности ячейки, однако он не позволяет учесть цикличность работы элементов и процессы восстановления информации.

Отметим, что наличие мажоритара на входе некоторой группы не связано с привязкой мажоритара к входу или выходу блока. В любом случае граф, описывающий структуру устройства, можно разбить на ячейки или домены для расчета надежности. В случае размещения мажоритаров на выходах групп узлов количество мажоритаров может быть несколько меньше, однако несколько групп резервированных узлов могут попасть в одну ячейку, что усложнит расчет.

Рассмотрим пример сети, приведенный на рис. 2. В ячейке *5* выходы узлов *b* не являются выходами ячейки, а используются как входы мажоритаров *c*. Выделение групп *c* и *d* в отдельную ячейку невозможно, поскольку узлы *d* связаны по входу непосредственно с мажоритарами *a*.

Наличие у ячейки «многослойной» структуры, когда она включает мажоритары, не являющиеся входными элементами ячейки, усложняет задачу анализа, так как не позволяет построить простой критерий допустимости заданной комбинации отказов.



■ ***Рис. 3.*** Преобразованная ячейка

Не изменяя структуры связей, можно рассматривать выходы узлов *b* как выходы ячейки, а входы узлов *c* — как входы ячейки. Тогда мы получим структуру, показанную на рис. 3.

С точки зрения расчета надежности узлы *f* и *g* можно рассматривать как один узел. В результате ячейка представляется «двухслойной» структурой: в первом слое стоят мажоритары, а во втором — резервированные узлы. Входы каждого узла подключены к одному или нескольким мажоритарам.

Любую ячейку, применяя описанное преобразование ко всем мажоритарам, не являющимся входными элементами, можно привести к «двухслойному» виду. Далее будем строить оценки надежности такой преобразованной ячейки.

Оценка надежности ячейки с учетом периода восстановления информации

Пусть M — множество троек узлов (modules) ячейки, $|M| = N_m$; V — множество троек мажоритаров (voters), $|V| = N_v$. К выходу каждого мажоритара подключен один или несколько узлов. Найдем функцию out:$V \to 2^M$, определяющую множество троек узлов, подключенных к выходам заданной тройки мажоритаров. Например, на рис. 2 out(a) = {b, d}.

Для построения оценки надежности ячейки с учетом процессов восстановления рассмотрим более подробно отказы в узлах и мажоритарах.

Пусть поток отказов каждого узла i и каждого мажоритара i — пуассоновский с интенсивностью λ_i и δ_i соответственно.

Пусть обновление информации в узле M_i выполняется периодически с периодом t_{Ri}. Будем рассматривать поведение ячейки на интервале времени T таком, что T кратно любому t_{Ri}, причем для $\forall i \in 1{:}N_v$, $\delta_i T << 1$ и $\forall i \in 1{:}N_m$, $\lambda_i T << 1$.

Отказы в узлах и мажоритарах по-разному влияют на работоспособность ячейки. Отказ в узле нарушает его работоспособность до конца текущего периода обновления. Ячейка (и вся система) выходит из строя, если на каком-либо периоде обновления произошел отказ в двух или более экземплярах какого-либо узла.

Сам по себе информационный отказ мажоритара не приводит к отказу ячейки, поскольку мажоритар не имеет элементов памяти, и состояние его выходов восстанавливается, как только рассасывается заряд, порожденный попаданием частицы. Однако отказ мажоритара вызывает отказ всех подключенных к нему узлов, причем отказ

в узле устраняется только с началом следующего цикла обновления соответствующего узла. Ячейка откажет, если отказы в мажоритарах (в одном или нескольких) произойдут на одном цикле обновления некоторого узла и выведут из строя хотя бы два его элемента.

Поскольку узлы имеют пуассоновский поток отказов, вероятность ровно одного отказа в элементе j за период T равна $\lambda_j T e^{-\lambda_j T} \approx \lambda_j T(1-\lambda_j T)$. Вероятность двух отказов равна $\frac{\lambda_j^2 T^2}{2} e^{-\lambda_j T} \approx \frac{\lambda_j^2 T^2}{2}(1-\lambda_j T)$. Будем рассматривать разложение функции вероятности безотказной работы ячейки по степеням T до второй степени. Из приведенных выражений видно, что для этого достаточно рассмотреть случаи возникновения нуля, одного и двух отказов в элементах.

Рассмотрим вероятность исправности ячейки как сумму вероятностей трех несовместных событий, когда произошло ни одного, один и два отказа: $P(T) = S_0 + S_1 + S_2$. Случаи трех и более отказов рассматривать не будем, поскольку это величины малого порядка, и их разложение в степенной ряд имеет слагаемые со степенями T не менее трех.

Вероятность того, что не отказал ни один элемент в тройке, равна

$$e^{-3\lambda_i T} = 1 - 3\lambda_j T + \frac{9\lambda_j^2 T^2}{2} + o(T^2).$$

Для нуля отказов имеем

$$S_0 = \prod_{j\in 1:N_m}\left(1-3\lambda_j T+\frac{9\lambda_j^2T^2}{2}\right)\prod_{j\in 1:N_v}\left(1-3\delta_j T+\frac{9\delta_j^2T^2}{2}\right)\approx$$

$$\approx 1-3\sum_{j\in 1:N_m}\lambda_j T+9\sum_{\substack{i,j\in 1:N_m\\ i<j}}\lambda_i\lambda_j T^2+9\sum_{i\in 1:N_m}\frac{\lambda_i^2T^2}{2}-$$

$$-3\sum_{j\in 1:N_v}\delta_j T+9\sum_{\substack{i,j\in 1:N_v\\ i<j}}\delta_i\delta_j T^2+$$

$$+9\sum_{i\in 1:N_v}\frac{\delta_i^2T^2}{2}+9\sum_{\substack{i\in 1:N_m,\\ j\in 1:N_v}}\lambda_i\delta_j T^2.$$

При ровно одном отказе в любом элементе система остается работоспособной.

Найдем вероятность ровно одного отказа в тройке узлов j:

$$3\lambda_j T e^{-\lambda_j T}\left(e^{-\lambda_j T}\right)^2 \approx 3\lambda_j T(1-3\lambda_j T).$$

Тогда вероятность одного отказа узла в ячейке (и отсутствия отказов в мажоритарах)

$$S_{1M} = \sum_{j\in 1:N_m}\left\{3\left(\lambda_j T-3\lambda_j^2T^2\right)\prod_{\substack{i\in 1:N_m\\ i\neq j}}\left[1-3\lambda_i T+\frac{9\lambda_i^2T^2}{2}\right]\right\}\times$$

$$\times\prod_{i\in 1:N_v}\left[1-3\delta_i T+\frac{9\delta_i^2T^2}{2}\right]\approx 3\sum_{j\in 1:N_m}\left(\lambda_j T-3\lambda_j^2T^2\right)\times$$

$$\times\left[1-3\sum_{\substack{i\in 1:N_m\\ i\neq j}}\lambda_i T+9\sum_{\substack{i,k\in 1:N_m\\ i,k\neq j,i<k}}\lambda_i\lambda_k T^2+9\sum_{\substack{i\in 1:N_m\\ i\neq j}}\frac{\lambda_i^2T^2}{2}\right]\times$$

$$\times\left(1-3\sum_{i\in 1:N_v}\delta_i T\right)\approx 3\sum_{j\in 1:N_m}\left(\lambda_j T-3\lambda_j^2T^2-3\sum_{\substack{i\in 1:N_m,\\ i\neq j}}\lambda_i\lambda_j T^2\right)\times$$

$$\times\left(1-3\sum_{i\in 1:N_v}\delta_i T\right)\approx 3\sum_{j\in 1:N_m}\lambda_j T-9\sum_{\substack{i,j\in 1:N_m\\ i\neq j}}\lambda_i\lambda_j T^2-$$

$$-9\sum_{i\in 1:N_m}\lambda_i^2T^2-9\sum_{\substack{i\in 1:N_v\\ j\in 1:N_m}}\delta_i\lambda_j T^2.$$

Аналогично вероятность одного отказа в мажоритаре

$$S_{1V} = 3\sum_{j\in 1:N_v}\delta_j T-9\sum_{\substack{i,j\in 1:N_v\\ i\neq j}}\delta_i\delta_j T^2-$$

$$-9\sum_{i\in 1:N_v}\delta_i^2T^2-9\sum_{\substack{i\in 1:N_v\\ j\in 1:N_m}}\delta_i\lambda_j T^2.$$

Вероятность возникновения ровно одного отказа $S_1 = S_{M1} + S_{V1}$.

Оценим теперь вероятность исправности ячейки при двух отказах:

$$S_2 = \sum_{j,k\in 1:N_e} P_{\text{отк}(j,k)}P_{\text{испр\_отк}(j,k)}P_{\text{испр\_ост}(j,k)},$$

где N_e — количество всех троек в ячейке ($N_e = N_m + N_v$); $P_{\text{отк}(j,\,k)}$ — вероятность возникновения по одному отказу в тройках j и k (если $j = k$ — то вероятность двух отказов в одной тройке); $P_{\text{испр\_отк}(j,\,k)}$ — апостериорная вероятность того, что система останется исправной после двух таких отказов; $P_{\text{испр\_ост}(j,\,k)}$ — вероятность того, что в остальных тройках отказов не будет. При разложении в ряд $P_{\text{испр\_ост}(j,\,k)}$ по степеням T получим $1 + k_1T + k_2T^2 + \ldots$. Поскольку мы рассматриваем разложение $P(T)$ до второй степени, данным множителем можно пренебречь, и далее он не указывается.

Будем отдельно рассматривать взаимоисключающие варианты, когда отказы возникли в различных тройках узлов, в одной тройке узлов,

в различных тройках мажоритаров, одной тройке мажоритаров, мажоритаре и узле:

$$S_2 = S_{M2} + S_{M1} + S_{V2} + S_{V1} + S_{VM} \approx$$
$$\approx \sum_{j,k\in 1:N_m} P_{\text{отк\_}M(j,k)} P_{\text{испр\_отк\_}M(j,k)} +$$
$$+ \sum_{j\in 1:N_m} P_{\text{отк\_}M(j,j)} P_{\text{испр\_отк\_}M(j,j)} +$$
$$+ \sum_{j,k\in 1:N_v} P_{\text{отк\_}V(j,k)} P_{\text{испр\_отк}V(j,k)} +$$
$$+ \sum_{j\in 1:N_m} P_{\text{отк\_}V(j,j)} P_{\text{испр\_отк\_}V(j,j)} +$$
$$+ \sum_{\substack{i\in 1:N_v \\ j\in 1:N_m}} P_{\text{отк\_}V_iM_j} P_{\text{испр\_отк\_}V_iM_j}.$$

В отличие от приведенных ранее рассуждений здесь необходимо отдельно рассматривать узлы и мажоритары.

Для случая одного отказа в двух различных тройках узлов (система при этом всегда остается исправной)

$$S_{M2} = \sum_{\substack{j,k\in N_m \\ j<k}} 3\lambda_j T e^{-3\lambda_j T} 3\lambda_k T e^{-3\lambda_k T} \cdot 1 \approx$$
$$\approx 9 \sum_{\substack{j,k\in N_m \\ j<k}} \lambda_j \lambda_k T^2.$$

Для случая двух отказов в одной тройке узлов важно, произошли ли отказы в одном экземпляре или в разных. Если отказы произошли в одном экземпляре, система останется исправной. Если отказы произошли в разных экземплярах на одном периоде обновления, система откажет, иначе останется исправной.

Найдем вероятность двух отказов в тройке:

$P_{j\ \text{разн}}$ — вероятность возникновения отказов в различных экземплярах узла;

$P_{j\ \text{одинак}}$ — вероятность возникновения двух отказов в одном экземпляре узла.

Вероятность двух отказов равна

$$P_{j,\text{разн}} + P_{j,\text{одинак}} = C_3^2 \left(\lambda_j T e^{-\lambda_j T}\right)^2 e^{-\lambda_j T} +$$
$$+ C_3^1 \frac{\lambda_j^2 T^2}{2} e^{-3\lambda_j T} \approx \frac{9}{2}\lambda_j^2 T^2.$$

$P_{1\text{гр}}$ — вероятность того, что два отказа произойдут в одном экземпляре при условии, что сбои произошли, равна 1/3;

$P_{2\text{гр}}$ — вероятность того, что два отказа произойдут в разных экземплярах при условии, что сбои произошли, равна 2/3;

$P_{\text{испр\_1гр}}$ — вероятность того, что система останется исправной при условии, что два отказа произойдут в одном экземпляре, равна 1;

$P_{\text{испр\_2гр}}$ — вероятность того, что система останется исправной при условии, что два отказа произойдут в разных экземплярах, равна вероятности того, что отказы не произойдут на одном цикле обновления, и составит $1 - \frac{t_{Rj}}{T}$.

$$S_{M1} = \sum_{j\in 1:N_m} (P_{j,\text{разн}} + P_{j,\text{одинак}}) \times$$
$$\times (P_{1\text{гр}} P_{\text{испр\_1гр}} + P_{2\text{гр}} P_{\text{испр\_2гр}});$$

$$S_{M1} = \sum_{j\in 1:N_m} \frac{9}{2}\lambda_j^2 T^2 \left(\frac{1}{3} + \frac{2}{3}\left(1 - \frac{t_{Rj}}{T}\right)\right) =$$
$$= \frac{9}{2} \sum_{j\in 1:N_m} \lambda_j^2 T^2 - 3 \sum_{j\in 1:N_m} \lambda_j^2 t_{Rj} T.$$

Вероятность отказа в двух различных мажоритарах j и k составляет $9\delta_j\delta_k T^2$.

Система откажет, если отказы произойдут на одном цикле обновления хотя бы для одного узла, подключенного к обоим мажоритарам, причем отказы произойдут в разных элементах тройки:

$$S_{V2} = \sum_{j,k\in 1:N_v} 9\delta_j\delta_k T^2 \cdot P_{\text{испр\_отк\_}V(j,k)}.$$

Можно показать, что вероятность появления отказов на одном цикле обновления составляет

$$\frac{K \cdot t_{R\_\max(j,k)}}{T},$$

где $1 \le K \le 2$, а $t_{R\_\max(j,k)} = \max\{t_{Ri}\}_{i\in \text{out}(j)\cap \text{out}(k)}$.

При кратных периодах обновления в узлах $K = 1$.

Если два мажоритара, в которых произошли отказы, не имеют общих выходных узлов, такое событие не может вызвать отказа ячейки. Учтем этот случай, определив $t_{R\_\max(j,k)} = 0$ при $\text{out}(j)\cap\text{out}(k) = \varnothing$.

Вероятность появления отказов в разных элементах при условии, что отказы возникли, составляет 2/3. Тогда вероятность исправности ячейки при двух отказах в различных тройках мажоритаров

$$S_{V2} = \sum_{j,k\in 1:N_v} 9\delta_j\delta_k T^2 \left(1 - \frac{2}{3}\frac{K_{jk} \cdot t_{R\_\max(j,k)}}{T}\right) =$$
$$= 9 \sum_{j,k\in 1:N_v} \delta_j\delta_k T^2 - 6 \sum_{j,k\in 1:N_v} \delta_j\delta_k K_{jk} \cdot t_{R\_\max(j,k)} T.$$

Вероятность возникновения двух сбоев в тройке мажоритаров j составляет $\frac{9}{2}\delta_j^2 T^2$. Аналогично предыдущему случаю, система откажет, если отказы произойдут на одном цикле обновления хотя бы для одного узла, подключенного к мажо-

ритару, причем отказы произойдут в разных элементах тройки:

$$S_{V1} = \sum_{j\in 1:N_v} \frac{9}{2}\delta_j^2 T^2 \cdot P_{\text{испр\_отк\_}V(j)} =$$

$$= \sum_{j\in 1:N_v} \frac{9}{2}\delta_j^2 T^2 \left(1 - \frac{2}{3}\frac{K_j \cdot t_{R\_\max(j)}}{T}\right) =$$

$$= \frac{9}{2}\sum_{j\in 1:N_v} \delta_j^2 T^2 - 3\sum_{j1:N_v} \delta_j^2 K \cdot t_{R\_\max(j)} T,$$

$$t_{R\_\max(j)} = \max\{t_{Ri}\}_{i\in \text{out}(j)}.$$

Осталось рассмотреть случай, когда возникает один сбой в мажоритаре и один в узле. Система откажет, если узел подключен к данному мажоритару и если отказы произойдут на одном цикле обновления, причем в разных элементах.

Определим $L(i, k)$ следующим образом:

$$L(i,k) = \begin{cases} 1, k \in \text{out}(i) \\ 0, k \notin \text{out}(i) \end{cases}.$$

Вероятность возникновения двух отказов в мажоритаре i и узле j

$$P_{\text{отк\_}V_iM_j} = 3\delta_i T e^{-3\delta_i T} 3\lambda_k T e^{-3\lambda_k T}.$$

Тогда

$$S_{VM} = \sum_{\substack{i\in 1:N_v \\ j\in 1:N_m}} P_{\text{отк\_}V_iM_j} P_{\text{испр\_отк\_}V_iM_j} =$$

$$= \sum_{\substack{i\in 1:N_v \\ k\in 1:N_m}} \left(3\delta_i T e^{-3\delta_i T} 3\lambda_k T e^{-3\lambda_k T}\left(1 - L(i,k)\cdot\frac{2}{3}\frac{t_{Ri}}{T}\right)\right) =$$

$$= 9\sum_{\substack{i\in 1:N_v \\ k\in 1:N_m}} \delta_i\lambda_k T^2 - 6\sum_{\substack{i\in 1:N_v \\ k\in 1:N_m}} L(i,k)\cdot\delta_i\lambda_k t_{Rk} T =$$

$$= 9\sum_{\substack{i\in 1:N_v \\ k\in 1:N_m}} \delta_i\lambda_k T^2 - 6\sum_{\substack{i\in 1:N_v \\ j\in \text{out}(i)}} \delta_i\lambda_j t_{Rj} T;$$

$$S_2 = 9\sum_{\substack{j,k\in N_m \\ j<k}} \lambda_j\lambda_k T^2 + \frac{9}{2}\sum_{j\in 1:N_m} \lambda_j^2 T^2 - 3\sum_{j\in 1:N_m} \lambda_j^2 t_{Rj} T +$$

$$+ 9\sum_{j,k\in 1:N_v} \delta_j\delta_k T^2 - 6\sum_{j,k\in 1:N_v} \delta_j\delta_k K_{jk}\cdot t_{R\_\max(j,k)} T +$$

$$+ \frac{9}{2}\sum_{j\in 1:N_v} \delta_j^2 T^2 - 3\sum_{j\in 1:N_v} \delta_j^2 K_j \cdot t_{R\_\max(j)} T +$$

$$+ 9\sum_{\substack{i\in 1:N_v \\ k\in 1:N_m}} \delta_i\lambda_k T^2 - 6\sum_{\substack{i\in 1:N_v \\ j\in \text{out}(i)}} \delta_i\lambda_j t_{Rj} T.$$

Построим теперь выражение для $P(T)$, суммируя выражения, полученные для S_0, S_1, S_2, и выполняя сокращения слагаемых:

$$P(T) = S_0 + S_1 + S_2 =$$

$$= 1 - 3\sum_{j\in 1:N_m} \lambda_j^2 t_{Rj} T - 6\sum_{j,k\in 1:N_v} \delta_j\delta_k K_{jk}\cdot t_{R\_\max(j,k)} T -$$

$$- 3\sum_{j\in 1:N_v} \delta_j^2 K_j \cdot t_{R\_\max(j)} T - 6\sum_{\substack{i\in 1:N_v \\ j\in \text{out}(i)}} \delta_i\lambda_j t_{Rj} T.$$

Получим теперь оценку вероятности исправного состояния системы через время T_M после старта. Пусть $T_M = NT$. Тогда $P(T_M) = P(T)^N = e^{\ln(P(T))N} \approx e^{(1-P(T))\frac{T_M}{T}}$.

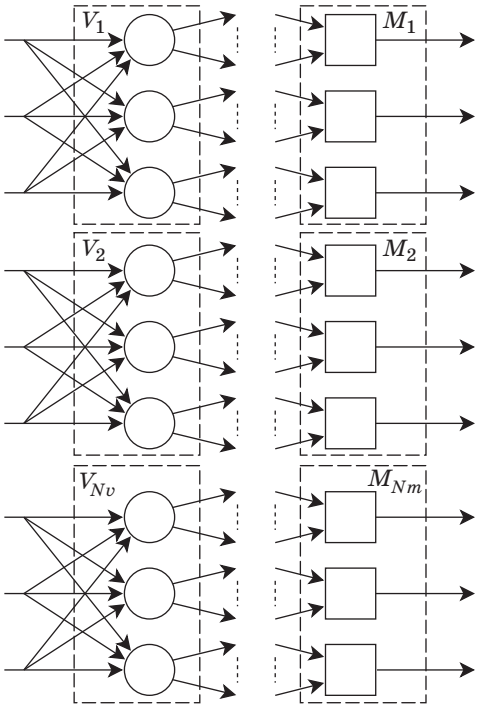
Подставляя $P(T)$, получаем, что $P(T_M) \approx e^{-\lambda_s T_M}$, где

$$\lambda_s = 3\sum_{j\in 1:N_m} \lambda_j^2 t_{Rj} + 3\sum_{j\in 1:N_v} \delta_j^2 K_j \cdot t_{R\_\max(j)} +$$

$$+ 6\sum_{j,k\in 1:N_v} \delta_j\delta_k K_{jk}\cdot t_{R\_\max(j,k)} + 6\sum_{\substack{i\in 1:N_v \\ j\in \text{out}(i)}} \delta_i\lambda_j t_{Rj}.$$

Таким образом, для произвольной ячейки с «двухслойной» структурой, показанной на рис. 4, получена оценка вероятности безотказной работы с учетом процессов восстановления информации в узлах. Из полученного выражения видно, что суммарная интенсивность потока отказов ячейки линейно зависит от периода восстановления информации.

Поскольку отказы в ячейках независимы, вероятность исправности системы составит $P_{\text{sys}}(T_M) = \prod_{i\in 1:N\text{cell}} P_i(T_M)$, где P_i — вероятность исправно-



■ ***Рис. 4.*** Двухслойная структура ячейки

сти *i*-й ячейки. Периоды восстановления узлов различных ячеек могут быть разными.

Заключение

Предложенный метод расчета надежности устройств учитывает периодичность процессов в резервированных узлах, а также структуру информационных связей, расположение и параметры надежности мажоритаров. Учет влияния мажоритаров на надежность устройства особенно важен при резервировании на уровне отдельных регистров или операционных узлов, когда сложность мажоритара сравнима со сложностью резервируемого узла, а количество мажоритаров велико. Метод может быть обобщен на узлы, защищенные не троированием, а помехоустойчивым кодированием информации.

Полученные результаты позволяют не только рассчитывать надежность, но и решать задачи синтеза. Можно планировать распределение «ненадежности» по узлам с учетом их уязвимости к внешним воздействиям и рассчитывать необходимые периоды восстановления. Это позволяет разработать методику проектирования устройств с учетом требований ко всем трем показателям: надежности, сложности реализации и быстродействию.

Литература

1. **Глухих М. И., Максименко С. Л., Мелехин В. Ф., Филиппов А. С.** Организация и проектирование высоконадежных вычислительных систем // Научно-технические ведомости СПбГПУ. 2011. № 6.1(138). С. 54–61.
2. **Максименко С. Л., Мелехин В. Ф., Филиппов А. С.** Анализ проблемы построения радиационно-стойких информационно-управляющих систем // Информационно-управляющие системы. 2012. № 2. С. 18–25.
3. **Максименко С. Л., Мелехин В. Ф.** Анализ надежности функциональных узлов цифровых СБИС со структурным резервированием и периодическим восстановлением работоспособного состояния // Информационно-управляющие системы. 2013. № 2. С. 18–23.
4. **Глухих М. И.** Разработка методов синтеза информационно-управляющих систем специального назначения со структурным резервированием: автореф. дис. ... канд. техн. наук. — СПб.: СПбГПУ, 2006. — 29 с.
5. **Jacob A. Abraham, Daniel P. Siewiorek.** An Algorithm for the Accurate Reliability Evaluation of Triple Modular Redundancy Networks // IEEE Transactions on Computers. 1974. Vol. C-23. N 7. P. 682–692.

УДК 658.562.3

МОДИФИКАЦИЯ АЛГОРИТМОВ УПРАВЛЕНИЯ, ИСПОЛЬЗУЮЩИХ ПРАВИЛА НЕЧЕТКОГО УСЛОВНОГО ВЫВОДА

В. Г. Чернов,
доктор экон. наук, профессор
Владимирский государственный университет
им. Александра Григорьевича и Николая Григорьевича Столетовых

Анализируются недостатки известных алгоритмов управления на основе правил нечеткого условного вывода. Предлагается модификация алгоритмов, устраняющая выявленные недостатки. Представлены результаты моделирования, показывающие, что новый подход позволяет получить лучшее качество управления.

Ключевые слова *— нечеткое множество, функция принадлежности, правила нечеткого условного вывода.*

Введение

Нечеткое управление различного рода техническими объектами в настоящее время получило достаточно широкое распространение, прежде всего, в тех приложениях, где применение методов классической теории автоматического управления осложнено трудностью создания адекватных математических описаний. Разработанные нечеткие контроллеры в ряде случаев обеспечивают качество управления, не уступающее классическим регуляторам [1–5].

Основные алгоритмы нечеткого управления

Среди известных алгоритмов нечеткого управления наиболее распространены алгоритмы Мамдани, Сукамото, Ларсена, которые основаны на правилах нечеткого условного вывода (ПНВ), образующих базу знаний нечеткого контроллера.

Независимо от вида алгоритма все они содержат следующие этапы:

1) фаззификацию, когда устанавливается соответствие между числовыми значениями входных переменных и определенными для них лингвистическими значениями;

2) поиск в базе знаний нечеткого контроллера подходящего правила или правил нечеткого управления;

3) обработку ПНВ, которая включает в себя свертку условий в условной части правил и формирование нечеткого вывода (вычисление импликации). Если при формировании управления предполагается использование нескольких правил, то формируется интегральный вывод;

4) дефаззификацию — преобразование нечеткого вывода в числовое значение для формирования управляющего воздействия.

Этапы 1, 2, 4 выполняются во всех алгоритмах аналогично, а различие между ними заключается лишь в формировании нечеткого вывода.

Поскольку именно третий этап является предметом рассмотрения настоящей работы, то остановимся на нем более подробно, взяв за основу алгоритм Мамдани. Отметим, что алгоритмы Сукамото и Ларсена отличаются от алгоритма Мамдани только в части формирования управляющего воздействия.

Рассмотрим упрощенный вариант, когда в управлении используются две входные переменные x и y, для которых определены лингвистические значения:

$L_x = \{A_i, i = 1, ..., N\}$, $L_y = \{B_j, j = 1, ..., M\}$, которым соответствуют нечеткие множества с функциями принадлежности $\mu_{A_i}(x)$, $\mu_{B_j}(y)$, $x \in U_x$, $y \in U_y$, где U_x, U_y — универсальные множества.

Положим для простоты $N = 4$, $M = 3$ (рис. 1). Нечеткий вывод представлен лингвистическими значениями $L_z = \{C_k, k = 1, ..., K\}$, $(K = 3)$.

Пусть управление осуществляется по правилам:

$$P1: \text{если } <x = A_2> \text{ и } <y = B_2>, \text{ то } <z = C_2>;$$

$$P2: \text{если } <x = A_3> \text{ и } <y = B_3>, \text{ то } <z = C_3>. \quad (1)$$

Согласно алгоритму Мамдани (аналогично для Сукамото и Ларсена), сначала выполняется свертка условий в левой части правил и вычисляются соответствующие функции принадлежности:

$$\mu_1 = \mu_{A_2}(x) \cap \mu_{B_2}(y);$$
$$\mu_2 = \mu_{A_3}(x) \cap \mu_{B_3}(y). \quad (2)$$

Сложность состоит в том, что операция пересечения определена для нечетких множеств, заданных на одном и том же универсальном множестве. В нашем случае это условие не выполняется. Поэтому в известных алгоритмах операция пересечения заменяется операцией min, которая выполняется не над соответствующими функциями принадлежности, а над их так называемыми синглтонами, т. е. значениями соответствующих функций принадлежности при конкретных значениях входных переменных, поступающих на вход системы в момент времени t (см. рис. 1).

На рисунке

$$\alpha_1 = \mu_{A_2}(x_t) \cap \mu_{B_2}(y_t) =$$
$$= \min[\mu_{A_2}(x_t), \mu_{B_2}(y_t)] = \min(\alpha_1, \beta_1); \quad (3)$$
$$\beta_2 = \mu_{A_3}(x_t) \cap \mu_{B_3}(y_t) =$$
$$= \min[\mu_{A_3}(x_t), \mu_{B_3}(y_t)] = \min(\alpha_2, \beta_2), \quad (4)$$

где x_t, y_t — текущие значения входных переменных.

На наш взгляд, в этом решении есть несколько недостатков.

По существу на данном наборе значений входных переменных многокритериальная задача сводится к однокритериальной. Правило $P1$ приводится к виду

$$P1: \text{если} <x = A_2>, \text{то} <z = C_2>, \text{а } P2 - P2:$$
$$\text{если} <y = B_3>, \text{то} <z = C_3>. \quad (5)$$

Очевидно, что это — другие правила и вообще другая ситуация управления.

Первое правило в некоторых пределах практически не реагирует на изменение второй входной переменной, пока $\alpha_1 \le \beta_1$. В какой-то мере ситуацию спасает наличие второго правила, но все равно это обстоятельство скажется на качестве управления. Аналогичные соображения можно высказать и в отношении правила $P2$ (5). Очевидно, что это же будет иметь место и при большем числе условий в левой части правил управления.

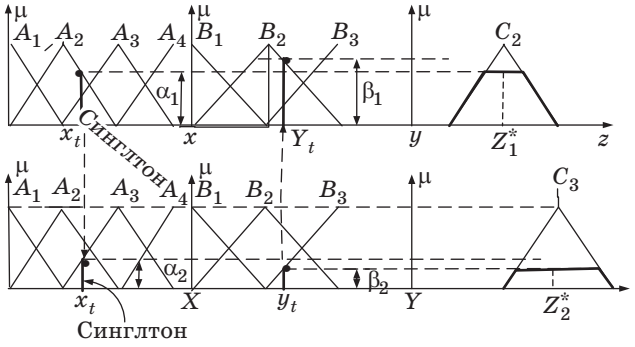
Кроме того, указанные выше алгоритмы не способны различить ситуации, представленные, например, наборами правил:

$$\text{если} <x = A_2> \text{ и } <y = B_2>, \text{то} <z = C_2>;$$
$$\text{если} <x = A_3>, \text{то если} <y = B_3>, \text{то} <z = C_3>; \quad (6)$$

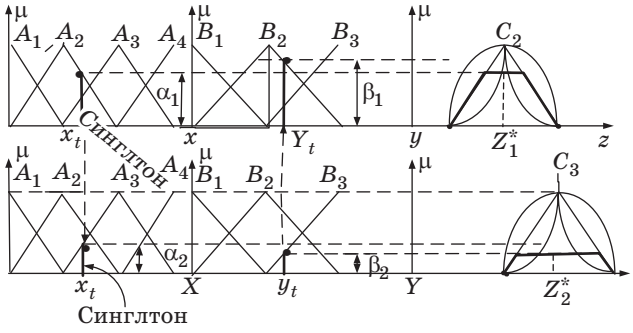
$$\text{если} <x = A_3> \text{ и } <y = B_3>, \text{то} <z = C_3>;$$
$$\text{если} <x = A_2>, \text{то если} <y = B_2>, \text{то} <z = C_2>. \quad (7)$$

Нетрудно показать, что в обеих ситуациях (6), (7) будет принято одно и то же решение. Это известный недостаток алгоритма Мамдани. Следует отметить, что правила вида (6), (7) использовались в работе самого Мамдани [5]. Необходимо также отметить, что в указанных алгоритмах в определенной мере теряется влияние вида функции принадлежности (рис. 2). Особенно это проявляется в алгоритме Сукамото, где итоговое решение вычисляется как взвешенное: $Z^* = (\alpha_1 z_1^* + \beta_2 z_2^*)/(\alpha_1 + \beta_2)$.

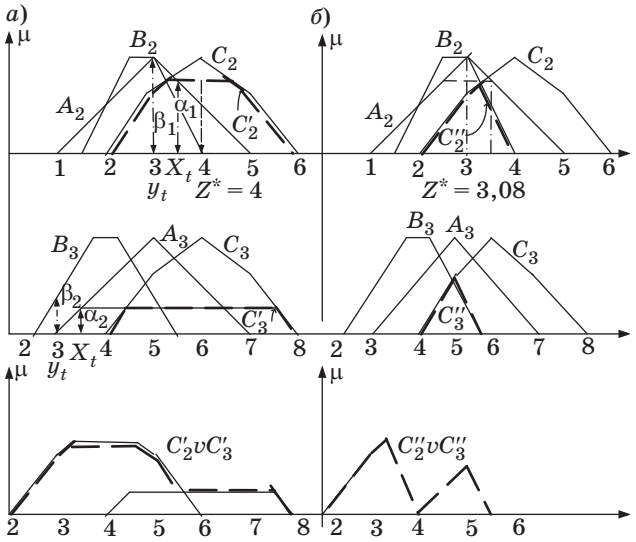
На рисунке, где показано три варианта функций принадлежности нечетких множеств, представляющих вывод, видно, что для любой симметричной функции принадлежности решение будет одним и тем же, так как значения z_1^* и z_2^* остаются неизменными независимо от вида функций принадлежности. Можно ожидать, что это отрицательно скажется на качестве управления. Для дополнительного доказательства этого положения рассмотрим ситуацию, когда все нечеткие множества, представляющие переменные ПНВ, определены на одном и том же универсальном множестве. На рис. 3, *а* представлен результат выполнения традиционного алгоритма Мамдани, когда операция пересечения заменена операцией min над синглтонами, на рис. 3, *б* — если ис-



■ ***Рис. 1.*** Обработка правил нечеткого вывода по алгоритму Мамдани



■ ***Рис. 2.*** Иллюстрация отсутствия влияния вида функции принадлежности на вывод в алгоритме Сукамото



■ ***Рис. 3.*** Нечеткий вывод: *а* — по традиционному алгоритму Мамдани с использованием синглтонов (α_1-, β_1-, α_2-, β_2-синглтоны); *б* — по алгоритму Мамдани с использованием операции пересечения (– – – — результат вывода)

пользуются пересечения нечетких множеств. Результат дефаззификации (решение или сигнал управления) для рис. 3, *а* — 4,58, для рис. 3, *б* — 3,2. Различия в результатах очевидны и составляют около 30 %. Конечно, это искусственная ситуация, но она нужна только как демонстрация возможных различий.

В замкнутых системах управления эти недостатки в какой-то мере устраняются именно за счет обратной связи, что позволяет строить нечеткие контроллеры. Однако в разомкнутых системах результаты не всегда получаются удовлетворительными. При этом надо отметить, что если для управления используется только одно правило, то указанные недостатки только усугубляются. Кроме того, применение этих алгоритмов в задачах многокритериального альтернативного выбора может дать противоречивые результаты [6].

Вернемся к алгоритму Мамдани. Ограничимся рассмотрением одного правила, например $P1$, результат обработки которого запишем следующим образом:

$$\mu_r = \mu_{C_2}(z) \cap [\mu_{A_2}(x) \cap \mu_{B_2}(y)]. \qquad (8)$$

Нетрудно показать, что это соотношение может быть представлено в виде

$$\mu_r = [\mu_{C_2}(z) \cap \mu_{A_2}(x)] \cap [\mu_{C_2}(z) \cap \mu_{B_2}(y)]. \qquad (9)$$

Соотношение (9) указывает на то, что нет необходимости выполнять свертку критериев в условной части ПНВ. Импликацию можно вычислять для каждого критерия, а свертку выполнять над частными импликациями. Таким образом можно обеспечить участие каждого критерия в формировании нечеткого управления. Конечно, пока остается нерешенной проблема вычисления пересечений множеств, стоящих в квадратных скобках соотношения (9).

Кроме простых правил вида (1) могут использоваться более сложные:

$$\text{если } <x = A> \text{ и } <y = B>,$$
$$\text{то если } <z = C> \text{ и } <q = D>, \text{ то } <p = H>; \qquad (10)$$

$$\text{если } <x = A> \text{ и } <y = B>,$$
$$\text{то } <z = C>, \text{ иначе } <z = D>. \qquad (11)$$

Согласно алгоритму Мамдани, для правила (10) может быть записано соотношение

$$\mu_1 = [(\mu_A \cap \mu_B)] \cap [(\mu_C \cap \mu_D) \cap \mu_H], \qquad (12)$$

которое можно переписать и так:

$$\mu_1 = \mu_A \cap [(\mu_C \cap \mu_H) \cap (\mu_D \cap \mu_H)] \cap \mu_B \cap [(\mu_C \cap \mu_H) \cap (\mu_D \cap \mu_H)],$$

т. е. все опять сводится к вычислению частных импликаций.

Правило (11) можно записать следующим образом [7]:

$$[(A \cap B) \to C] \cap [\overline{(A \cap B)} \to D].$$

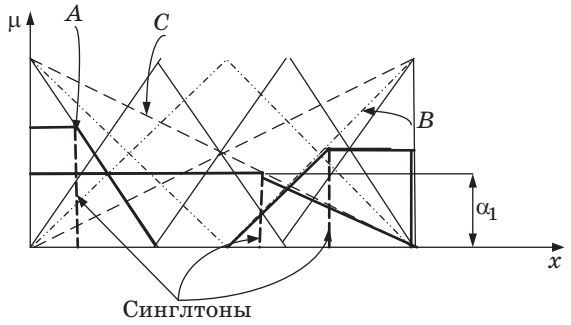
Учитывая доказанные соотношения (6), (7), получим

$$[(A \cap B) \to C] \cap [\overline{(A \cap B)} \to D] =$$
$$= [(A \cap C) \cap (B \cap C)] \cap [(\bar{A} \cup \bar{B}) \cap D] =$$
$$= [(A \cap C) \cap (B \cap C)] \cap [(\bar{A} \cap D) \cup (\bar{B} \cap D)],$$

что также подтверждает возможность отказа от свертки критериев в условной части правил вида (10) и (11).

Кроме этого, следует рассмотреть еще одну ситуацию, которая может возникнуть при обработке ПНВ (рис. 4).

Пусть имеется правило: если $<x = A>$ и $<y = B>$ и $<z = C>$, то $<h = D>$, при этом все нечеткие мно-



■ ***Рис. 4.*** Иллюстрация возможности возникновения пустого множества при свертке условий в ПНВ (функции принадлежности, полученные в результате фаззификации)

жества в условной части правила определены на одном и том же универсальном множестве.

На рисунке нетрудно видеть, что попытка построить свертку условий на основе операции пересечения приведет к получению пустого множества. Поэтому в известных алгоритмах операция пересечения опять же заменяется на операцию min, выполняемую не над нечеткими множествами, а над синглтонами. Результатом этой операции будет значение α_1. Недостатки этого подхода мы уже обсуждали ранее. Переход на вычисление частных импликаций позволяет предложить следующее решение. Построим из элементов условной части ПНВ группы, включающие все возможные комбинации нечетких множеств, дающие непустые пересечения (см. рис. 4) $(A\cap C)\neq 0$, $(B\cap C)\neq 0$, и вычислим частные импликации для этих групп $(A\cap C)\cap D$, $(B\cap C)\cap D$.

При обработке ПНВ кроме импликации Мамдани часто используется импликация Лукасевича [8], которая, например, для правила $P1$ запишется в виде

$$m_1 = 1\cap(1-\mu_A\cap\mu_B+\mu_C) = 1\cap(\overline{\mu_A\cap\mu_B}+\mu_C) = \\ = 1\cap(\overline{\mu}_A\cup\overline{\mu}_B+\mu_C) = \overline{\mu}_A\cup\overline{\mu}_B+\mu_C. \qquad (13)$$

Действуя аналогично, можно доказать справедливость равенств

$$m_1 = 1\cap(1-\mu_A+\mu_c) = \overline{\mu}_A+\mu_C,$$
$$m_2 = 1\cap(1-\mu_B+\mu_C) = \overline{\mu}_B+\mu_C.$$

Можно показать, что

$$m_1\cup m_2 = (\overline{\mu}_A+\mu_C)\cup(\overline{\mu}_B+\mu_C) = \overline{\mu}_A\cup\overline{\mu}_B+\mu_C,$$

т. е. объединение импликаций Лукасевича, вычисленных отдельно для нечетких множеств A и B, совпадает с импликацией, вычисленной по соотношению (11), что также указывает на то, что в случае применения импликации Лукасевича нет необходимости выполнять свертку критериев в условной части ПНВ.

Импликация Заде для простого правила «если A, то B» вычисляется по формуле $\max[\min(\mu_A, \mu_B), 1-\mu_A]$. Для правила $P1$ (1) импликация Заде будет вычисляться по формуле $\max[\min(\mu_{A\cap B}, \mu_C), 1-\mu_{A\cap B}]$. Запишем эту формулу несколько иначе:

$$\max[\min(\mu_{A\cap B},\mu_C),1-\mu_{A\cap B}] = \\ = [(\mu_A\cap\mu_B)\cap\mu_C]\cup(1-\mu_{A\cap B}) = [(\mu_A\cap\mu_B)\cap\mu_C]\cup \\ \cup\overline{\mu}_{A\cap B} = [(\mu_A\cap\mu_B)\cap\mu_C]\cup\overline{\mu_A\cap\mu_B} = \\ = [(\mu_A\cap\mu_C)\cap(\mu_B\cap\mu_C)]\cap(\overline{\mu}_A\cup\overline{\mu}_B) = \\ = (\overline{\mu}_A\cup\overline{\mu}_B)\cup[(\mu_A\cap\mu_C)\cap(\mu_B\cap\mu_C] = \\ = \overline{\mu}_A\cup[(\mu_A\cap\mu_C)\cap(\mu_B\cap\mu_C)]\cup\overline{\mu}_B\cup[(\mu_A\cap \\ \cap\mu_C)\cap(\mu_B\cap\mu_C)].$$

И, наконец, для классической нечеткой импликации для правила $P1\max[1-\mu_{A\cap B}, \mu_C]$ при $\mu_{A\cap B}\geq\mu_C$ получим

$$\max[1-\mu_{A\cap B},\mu_C] = \\ = [(1-\mu_A\cap\mu_B)\cup\mu_C] = [\overline{\mu_A\cap\mu_B}\cup\mu_C] = \\ = \overline{\mu}_A\cup\overline{\mu}_B\cup\mu_C = (\overline{\mu}_A\cup\mu_c)\cup(\overline{\mu}_B\cup\mu_C)).$$

Проведенный выше анализ показывает, что для наиболее распространенных вариантов обработки ПНВ можно отказаться от выполнения свертки критериев в условной части, выполняя расчет частных импликаций для каждого из критериев, и затем выполнять необходимые преобразования над этими частными импликациями. На наш взгляд, в этом есть одно существенное достоинство, состоящее в том, что здесь сохраняется участие всех критериев условной части ПНВ в его обработке. Определенным недостатком является то, что процедура обработки ПНВ получается более громоздкой, что, конечно, потребует больших вычислительных затрат. Однако современная элементная база не критична к подобного рода усложнениям. В случае программной реализации этот недостаток вообще малосущественен.

Отметим, что полученные результаты не являются окончательным решением задачи, а только обеспечивают возможности для дальнейшей модификации существующих алгоритмов обработки ПНВ.

Вернемся к одному из правил вида (1), например:

$P1$: если $<x = A_1>$ и $<y = B_1>$, то $<z = C_1>$.

Как было показано выше, результатом его обработки будет нечеткое множество с функцией принадлежности

$$\mu_1 = [\mu_{C1}(z)\cap\mu_{A1}(x)]\cap[\mu_{C1}(z)\cap\mu_{B1}(y)]. \qquad (14)$$

Поскольку нечеткие множества A_1, B_1, C_1 принадлежат различным универсальным множествам, то вычисление соответствующих пересечений в соотношении (14) по определению невозможно. Известно использование цилиндрических продолжений [7] для вычисления конъюнктивных форм, но для управления техническими объектами оно достаточно сложное.

В работе [6] была предложена новая операция над нечеткими множествами, которая первоначально была названа геометрической проекцией нечетких множеств. Это название было признано неудачным из-за близости по названию с операцией проекции нечетких множеств, поэтому было предложено новое название «тень нечетких множеств» $\mathrm{Sh}(A, B)$. Определим эту операцию следующим образом.

Тень нечеткого множества $\tilde{A}$ на нечеткое множество $\tilde{B}$ должна удовлетворять следующим условиям:

1) $\mathrm{Sh}(\tilde{A}, \tilde{B})$ — нечеткое множество;

2) $\mathrm{Sh}(\tilde{A}, \tilde{A}) = \tilde{A}$;

3) $\mathrm{Sh}(\tilde{A}, \tilde{B}) = \varnothing$, если хотя бы одно из множеств $\tilde{A}$ или $\tilde{B}$ пустое или множества $\tilde{A}$ и $\tilde{B}$ ортогональны.

Процедуру построения тени нечеткого множества $\tilde{A}$ на нечеткое множество $\tilde{B}$ определим следующим образом (рис. 5):

$$\mathrm{Sh}_{\varphi}(\tilde{A}, \tilde{B}) = \{\varphi[\mu_{\tilde{A}}(y), \mu_{\tilde{B}}(x)] / [y, x' = f(y)]\}, \quad (15)$$

где $f(y) = \dfrac{CG[\mu_{\tilde{B}}(x)]}{CG[\mu_{\tilde{A}}(y)]} y$ — проекционная функция; $CG[\mu_{\tilde{B}}(x)], CG[\mu_{\tilde{A}}(y)]$ — координаты центров тяжести фигур, ограниченных функциями принадлежности $\mu_{\tilde{A}}(y), \mu_{\tilde{B}}(x)$; φ — функционал, задающий вид преобразований над функциями принадлежности.

Смысл этой операции состоит в том, что в зависимости от взаимного расположения нечетких множеств и, соответственно, угла наклона проекционной прямой изменяется «тень» одного нечеткого множества, накладываемая на другое. Этим будет определяться степень взаимодействия оценок понятий, представляемых нечеткими множествами.

Нечеткое множество $\tilde{A}$, которое проецируется на другое нечеткое множество $\tilde{B}$, назовем источником тени. Нечеткое множество $\tilde{B}$, на которое проецируется тень нечеткого множества $\tilde{A}$, назовем приемником тени.

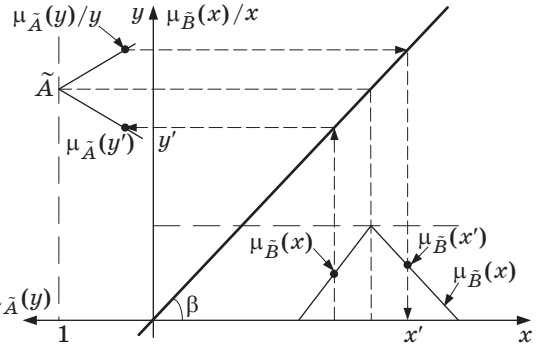
Тени типа min и max будут иметь место, если $\varphi = \min$ и $\varphi = \max$ соответственно:

$$\mathrm{Sh}_{\min}(\tilde{A}, \tilde{B}) = \{\min[\mu_{\tilde{A}}(y), \mu_{\tilde{B}}(x')] / y, x' = f(y)\}; \quad (16)$$

$$\mathrm{Sh}_{\max}(\tilde{A}, \tilde{B}) = \{\max[\mu_{\tilde{A}}(y), \mu_{\tilde{B}}(x')] / y, x' = f(y)\}. \quad (17)$$

Обратная тень

$$\mathrm{Sh}_{\varphi}^{-1}(\tilde{B}, \tilde{A}) = \{\varphi[\mu_{\tilde{B}}(x), \mu_{\tilde{A}}(y')] / x, y' = f(x)\}.$$



■ ***Рис. 5.*** Геометрическое представление операции «тень нечеткого множества»

Более подробно свойства этой операции рассмотрены в работах [6, 9].

В случае использования этой операции для управления техническими объектами применительно к алгоритмам Мамдани и Сукамото она была несколько видоизменена в части определения проекционной функции $f(y) = \dfrac{CG[\mu_{\tilde{B}}(x)]}{y_t} y$, где y_t — текущее значение переменной y (рис. 6).

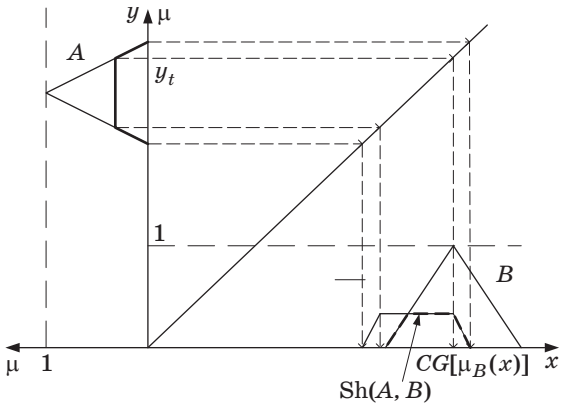
Для нечеткого правила условного вывода вида (1)

если $<x = A>$ и $<y = B>$ то $<z = C>$,

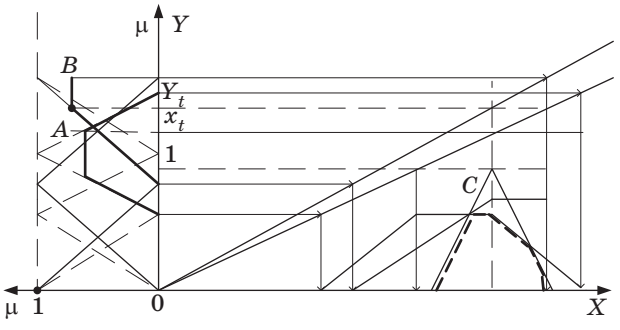
получим $\mu_1 = \mathrm{Sh}_{\varphi}(A, C) \cap \mathrm{Sh}_{\varphi}(B, C)$.

Графическая иллюстрация вывода представлена на рис. 7. Нетрудно видеть, что в данном случае сохраняется влияние каждого из условий на результирующий вывод и, кроме того, будет учтен характер используемых функций принадлежности.

И наконец, следует отметить, что предложенный метод обработки импликации свободен от



■ ***Рис. 6.*** Реализация операции $\mathrm{Sh}(A, B)$ с учетом результата фаззификации входной переменной (——— функция принадлежности, полученная после фаззификации переменной $y = y_t$; – – – — функция принадлежности нечеткого множества, представляющего $\mathrm{Sh}(A, B)$)



■ ***Рис. 7.*** Вывод с использованием операции «тень нечетких множеств» (– – – — результирующее нечеткое множество)

еще одного недостатка импликации Мамдани, состоящего в том, один и тот же результат получается для различных по логике условных правил. Как уже отмечалось, для правил если $<x = A>$ и $<y = B>$, то $<z = C>$ и если $<x = A>$, то если $<y = B>$, то $<z = C>$, будет получен один и тот же результат.

Если реализовать эти правила через операцию Sh, то для первого правила получим $\mathrm{Sh}_\varphi(B, C) \cap \cap \mathrm{Sh}_\varphi(A, C)$, для второго — $\mathrm{Sh}_\varphi[A, \mathrm{Sh}_\varphi(B, C)]$.

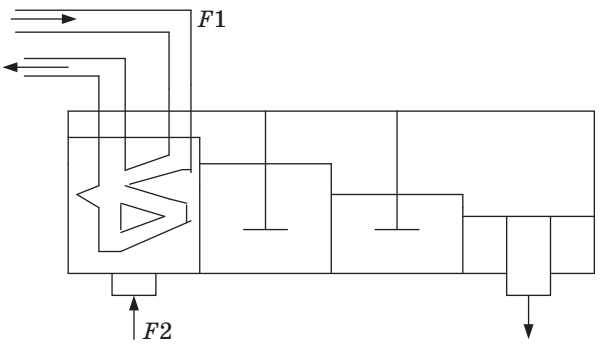
Нетрудно показать, что $\mathrm{Sh}_\varphi(B, C) \cap \mathrm{Sh}_\varphi(A, C) \neq \neq \mathrm{Sh}_\varphi[A, \mathrm{Sh}_\varphi(B, C)]$.

Рассмотрим возможности и результаты практического применения описанных выше подходов к решению задачи управления техническими объектами. В качестве объекта управления для проведения исследования была выбрана установка для нагрева жидкостей, так как объект данного типа является классическим примером САУ в теории управления, а также идеологически соответствует принципам нечеткого управления (сложная природа процессов и в то же время простая и понятная логика управления объектом) [10]. Упрощенная схема САУ представлена на рис. 8.

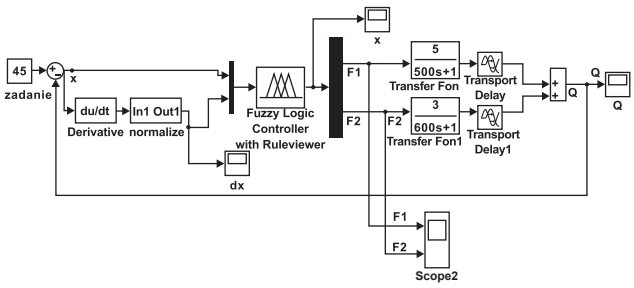
Обоснованием этого варианта является наличие, во-первых, стратегий управления, доказавших свою работоспособность, во-вторых, апробированной модели, реализованной в среде MatLab [11].

Бак с теплой водой разделяется на несколько отсеков, переменный поток холодной воды $F2$ проходит последовательно отсеки и покидает бак в последнем отсеке. Холодная вода нагревается в теплообменнике, в котором течет по трубам переменный поток горячей воды $F1$ с температурой 90 °С. Задача состоит в поддержании постоянной температуры воды в одном из отсеков и, по возможности, в сохранении постоянства потока $F2$ посредством регулирования динамических значений $F1$ и $F2$.

В среде Simulink программного комплекса MatLab построена модель (рис. 9) в соответствии со структурой, представленной в работе [10], где объект управления описывался соотношени-



■ ***Рис. 8.*** Схема САУ



■ ***Рис. 9.*** Модель САУ

ем $Q(s) = \alpha_1 H_1(s) F_1 - \alpha_2 H_2(s) F_2$, где α_1, α_2 — весовые коэффициенты;

$$H_1(s) = \frac{k_1 \exp(-\tau_1 s)}{T_1 s + 1},\ H_2(s) = \frac{k_2 \exp(-\tau_2 s)}{T_2 s + 1}.$$

Исходные данные для создания нечеткого регулятора были взяты из работы [10] с целью сопоставлять получаемые результаты более обоснованно.

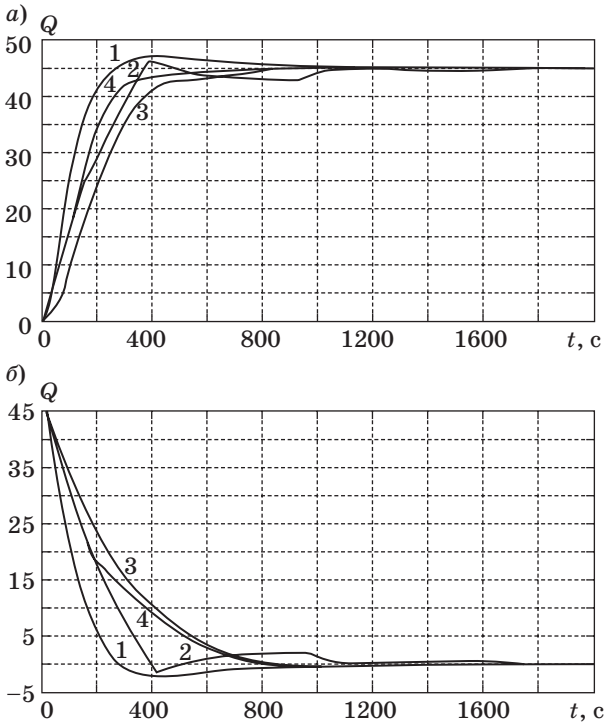
Диапазон изменения величин:

— поток горячей воды $F1 = 0{,}5 \div 42$;

— поток холодной воды $F2 = 1 \div 18$;

— заданное значение температуры нагрева +45 °С.

Результаты сравнительного моделирования стандартного и модифицированного алгоритмов Мамдани представлены на рис. 10, *а*, *б* и в таблице.



■ ***Рис. 10.*** Переходные процессы (*а*) и временные диаграммы изменения ошибки (*б*) в системе: *1*, *2* — стандартный и модифицированный алгоритмы Мамдани; *3*, *4* — стандартный и модифицированный алгоритмы Сукамото

■ Результаты сравнительного моделирования стандартных и модифицированных алгоритмов Мамдани и Сукамото

Параметр	Стандартный алгоритм		Модифицированный алгоритм	
	Мамдани	Сукамото	Мамдани	Сукамото
Время регулирования, с (точность 5 %)	1660	410	810	340
Перерегулирование, %	5,5	0	3,4	0

Интегральная оценка ошибки:
стандартный алгоритм Мамдани:

$$I_1 = \int_0^{2000} |Q| \mathrm{d}t = 8423(^\circ\mathrm{C} \cdot t);$$

модифицированный алгоритм:

$$I_2 = \int_0^{2000} |Q| \mathrm{d}t = 5912(^\circ\mathrm{C} \cdot t);$$

стандартный алгоритм Сукамото:

$$I_1 = \int_0^{1000} |Q| \mathrm{d}t = 1882(^\circ\mathrm{C} \cdot t);$$

модифицированный алгоритм:

$$I_2 = \int_0^{1000} |Q| \mathrm{d}t = 1790(^\circ\mathrm{C} \cdot t).$$

Нетрудно видеть, что модифицированный алгоритм дает заметно лучшее качество управления.

Заключение

Предложенная в статье модификация алгоритмов управления на основе ПНВ обеспечивает более высокое качество управления, чем традиционные алгоритмы. Это объясняется тем, что в процессе управления сохраняется влияние всех управляющих переменных, входящих в условную часть правил управления.

Литература

1. **Захаров В. Н., Ульянов С. В.** Нечеткие модели интеллектуальных промышленных регуляторов и системы управления // Изв. АН СССР. Техническая кибернетика. 1993. № 4. С. 189–205.
2. **Кузьмин В. Б., Травкин С. И.** Теория нечетких множеств в задачах управления и принципах устройства нечетких процессоров // Изв. АН СССР. Техническая кибернетика. 1992. № 5. С. 171–197.
3. **Асаи К., Вамада Д., Иваи С.** и др. Прикладные нечеткие системы: пер. с япон. / под ред. К. Тэрано, К. Асаи, М. Сугэно. — М.: Мир, 1993. — 368 с.
4. **Бураков М. В., Коновалов А. С.** Синтез нечетких логических регуляторов // Информационно-управляющие системы. 2011. № 1(50). С. 22–27.
5. **Mamdani E. H.** Application of fuzzy logic to approximate reasoning using linguistic systems // IEEE Trans. Comput. 1977. C-26. P. 1182–1191.
6. **Чернов В. Г.** Решение задач многокритериального альтернативного выбора на основе геометрической проекции нечетких множеств // Информационно-управляющие системы. 2007. № 1(26). С. 46–52.
7. **Малышев Н. Г., Берштейн Л. С., Боженюк А. В.** Нечеткие модели для экспертных систем в САПР. — М.: Энергоатомиздат, 1991. — 136 с.
8. **Борисов А. Н., Крумберг О. А., Федоров И. П.** Принятие решений на основе нечетких моделей: Примеры использования. — Рига: Зинатне, 1990. — 184 с.
9. **Чернов В. Г.** Нечеткие деревья решений (нечеткие позиционные игры) // Информационно-управляющие системы. 2010. № 5(48). С. 8–14.
10. **Kicker W. J. M., Van Nauta Lemke H. R.** Application of a fuzzy controller in a warm water plant // Automatica. 1976. Vol. 12. P. 301–308.
11. **Леоненков А. В.** Нечеткое моделирование в среде MATLAB и fuzzyTECH. — СПб.: БХВ-Петербург, 2005. — 736 с.

УДК 658.512.22

АЛГОРИТМИЗАЦИЯ ОБРАБОТКИ И ПЕРЕДАЧИ МЕТЕОРОЛОГИЧЕСКИХ ДАННЫХ В ЗАМКНУТОЙ СИСТЕМЕ УПРАВЛЕНИЯ «ПРИРОДА-ТЕХНОГЕНИКА»

Р. И. Сольницев,
доктор техн. наук, профессор
Санкт-Петербургский государственный университет аэрокосмического приборостроения
До Суан Чо,
аспирант
Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»

Излагается дальнейшее развитие теории и практики создания замкнутой системы управления «Природа-техногеника», предназначенной для эффективного снижения загрязняющих веществ, выбрасываемых промышленными предприятиями в атмосферу. Представлены алгоритмы обработки и передачи метеорологических данных в систему управления «Природа-техногеника».

Ключевые слова *— экология, загрязняющие вещества, система автоматического управления, метеорологическое обеспечение, алгоритм.*

Введение

Замкнутая система управления «Природа-техногеника» (ЗСУПТ) предназначена для минимизации загрязняющих веществ (ЗВ) в окружающей среде. Концепция ЗСУПТ и ее развитие изложены в предыдущих работах (например, [1]). В данной статье рассматривается ЗСУПТ в атмосфере. Важной задачей в процессе создания этой системы является разработка средств ее метеорологической поддержки в реальном времени. Поскольку проектирование ЗСУПТ проводится с помощью САПР, рассматриваются постановка задачи, алгоритмизация обработки, ввода метеорологической информации, а также некоторые результаты применения соответствующей подсистемы САПР ЗСУПТ.

Влияние метеорологических данных на ЗСУПТ

Скорость потока переноса, $\tilde{V}$, ЗВ в атмосфере от предприятий — источников ЗВ до датчиков измерения (рис. 1) существенно зависит от метеорологических данных в заданной окрестности источника ЗВ (природопользовательской зоне).

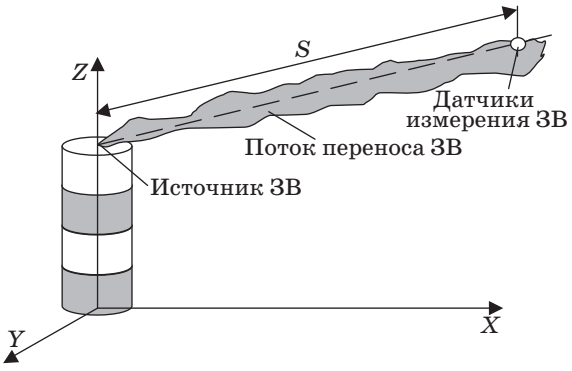
Основными составляющими метеорологического влияния на скорость потока переноса ЗВ $\tilde{V}$ являются скорость ветра V, температура T, давление P, влажность W.

Скорость потока $\tilde{V}$ определяет время запаздывания τ переноса ЗВ от источника до датчиков измерения концентраций ЗВ. ЗСУПТ представляет собой систему автоматического управления (САУ), в которой объект управления — распределенная система [2]. Структурная схема ЗСУПТ (рис. 2) включает устройства как отдельные звенья этой САУ [3].

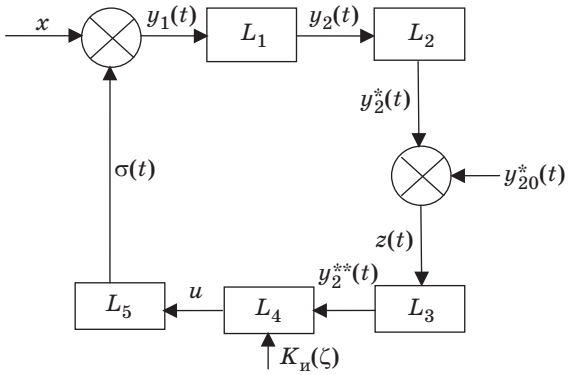
На структурной схеме обозначены операторы:

L_1 — переноса ЗВ от источника ЗВ до точки измерения параметров;

L_2 — измерительного устройства (датчика);



■ ***Рис. 1.*** Поток переноса ЗВ в атмосфере



■ ***Рис. 2.*** Структурная схема ЗСУПТ как САУ

L_3 — устройства преобразования данных измерения;

L_4 — устройства управления;

L_5 — агрегата очистки совместно с исполнительным устройством, а также сигналы:

x — возмущающее воздействие (компенсируемая составляющая топливных газов);

y_1 — рассогласование;

y_2 — измеряемая величина концентрации ЗВ;

y_2^* — результат измерения параметров;

y_{20}^* — допустимая величина концентрации ЗВ (y_2^*);

z — величина отклонения;

u — управление;

σ — компенсация возмущения;

y_2^{**} — преобразованный сигнал.

При известном расстоянии от источника ЗВ до датчика (S на рис. 1) скорость переноса $\tilde{V}$ является неизвестной функцией не только ветра (V), но и всех остальных метеорологических параметров.

Поскольку направление ветра определяется положением подвижного носителя датчиков измерения [3], то в дальнейшем рассматривается только составляющая скорости ветра по оси факела.

Методики расчетов концентраций ЗВ, усредненных за большие сроки на основе многолетних экспериментальных данных, хорошо известны и представлены в стандартных формах. Однако применение этих методик для управления очистными агрегатами в ЗСУПТ невозможно, поскольку в ЗСУПТ требуется непрерывное (или с приемлемой дискретностью) поступление управляющих сигналов на исполнительное устройство.

В структурной схеме ЗСУПТ регулятор представлен оператором L_4. В результате синтеза закона регулирования получен [3] пропорционально-интегрально-дифференциальный (ПИД) регулятор, и оператор L_4 имеет вид

$$w_4(p) = K_п + K_д p + K_и \frac{1}{p}, \qquad (1)$$

где $p = \frac{d}{dt}$.

Особенностью такой САУ является то, что объект управления представляет собой распределенную систему, математическая модель которой после приведения ее к форме «вход — выход» (оператор L_1 на рис. 2) имеет вид $w_1(p) = \frac{k_1}{T_1 p + 1} e^{-p\tau}$, где τ — запаздывание при переносе ЗВ от источника до датчика измерительной системы ЗСУПТ [2]. При этом требуемые характеристики процессов регулирования достигаются при соотношении $K_и \tau = a \cong 0{,}8 \div 1{,}2$, где $K_и$ — коэффициент передачи интегрирующего звена в ПИД-регуляторе [3].

С другой стороны, величина τ определяется скоростью переноса ЗВ $\tilde{V}$ и является функцией

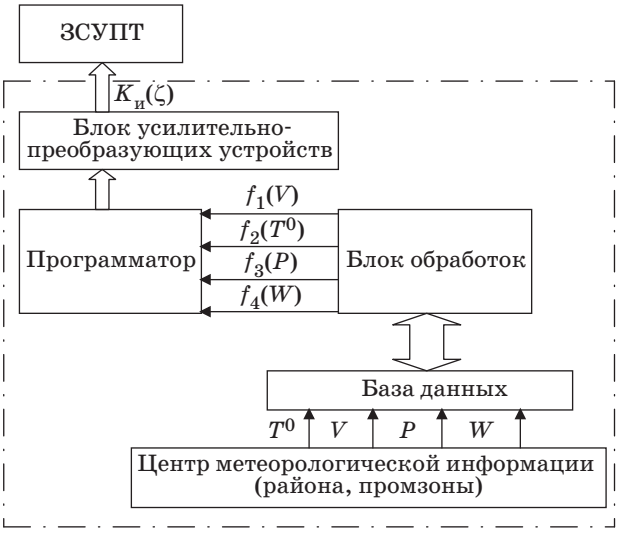
$$\tau = F(V, T^0, P, W, S), \qquad (2)$$

где величина расстояния S считается заданной.

В этой связи требуются автоподстройка коэффициента $K_и$ в ПИД-регуляторе при непрерывном (или с допустимой дискретностью) поступлении метеорологических данных и определение алгоритма автоподстройки с дальнейшим построением соответствующей программы контроллера, реализующего зависимость

$$K_и = F^0(V, T^0, P, W, S) = K_и(\zeta). \qquad (3)$$

Обработка и передача метеорологической информации в ЗСУПТ, в свою очередь, требуют разработки алгоритмов определения метеорологических характеристик: $f_1(V)$, $f_2(T^0)$, $f_3(P)$, $f_4(W)$ — и отношений между этими функциями (рис. 3). Переменные V, T^0, P, W в общем случае являются нестационарными случайными процессами (НСП). Функции f_1, f_2, f_3, f_4 рассматриваются в дальнейшем [см. (4)–(9)] как оценки НСП. На основе этих функций строятся массивы данных, поступа-



■ ***Рис. 3.*** Схема сбора, обработки и передачи метеорологической информации в ЗСУПТ

ющих в ЗСУПТ непрерывно или с допустимой дискретностью.

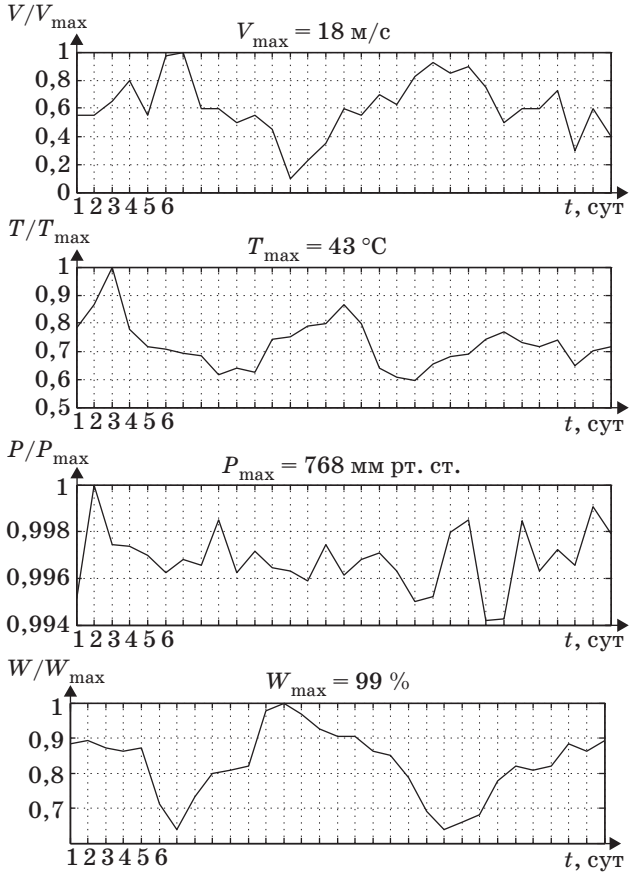
В соответствии с отмеченными обстоятельствами в статье рассматриваются:

1) алгоритмы обработки метеорологической информации V, T^0, P, W в виде математических ожиданий $M(V)$, $M(T^0)$, $M(P)$, $M(W)$, автокорреляционных и взаимно корреляционных функций R_{VV}, R_{TT}, R_{PP}, R_{WW}, R_{VT}, R_{VP}, R_{VW};

2) алгоритмизация управления автоподстройкой ПИД-регулятора.

Алгоритмы обработки метеорологических данных

В настоящей работе исходной является метеорологическая информация V, T^0, P, W, полученная из Центра метеорологической информации города Ханоя (Вьетнам) для микрорайона (промзоны) «Шок Шон» с долготой 106°03′ и широтой 20°39′ [4]. На рис. 4 представлены примеры реализаций, иллюстрирующие массив этой информации. Для обработки массива реализаций как НСП рассматриваются алгоритмы сглаживания [5]. Алгоритмы сглаживания использованы [6] для приведения НСП к эквивалентным стационарным с соответствующими оценками точности такого приведения. Алгоритмы обработки на основе сглаживания приводятся далее в виде оценок НСП.



■ ***Рис. 4.*** Примеры исходных реализаций метеорологической информации (январь 2011 г.)

Математическое ожидание

$$M\left(x_i(t)\right)=\frac{1}{T}\int\limits_{t-\frac{T}{2}}^{t+\frac{T}{2}} x_i(t)\mathrm{d}t,\ i=1,2,3,4, \qquad (4)$$

где T — интервал текущего сглаживания; $x_i(t)$ — исходные реализации, соответствующие V, T^0, P, W.

Автокорреляционная функция

$$R^i_{xx}(t,\lambda)=\frac{1}{T_1}\int\limits_{t-\frac{T_1}{2}}^{t+\frac{T_1}{2}} x^0_i(t)x^0_i(t+\lambda)\mathrm{d}t, \qquad (5)$$

где $\lambda=\mu\Delta\lambda$ — время задержки, $\mu=1, 2, 3, \ldots, n$, $\Delta\lambda$ — шаг квантования по λ; T_1 — интервал текущего сглаживания; $x^0_i(t)=x_i(t)-M\left(x_i(t)\right)$.

Средняя автокорреляционная функция по одной реализации определяется по формуле

$$R^i_{xx_{\text{ср}}}(\lambda)=\frac{1}{T_2}\int\limits_0^{T_2} R^i_{xx}(t,\lambda)\mathrm{d}t, \qquad (6)$$

где T_2 — время усреднения корреляционной функции $R^i_{xx}(t,\lambda)$ с численной оценкой близости обрабатываемого НСП к эквивалентному стационарному [6].

Средняя автокорреляционная функция по всем реализациям ансамбля будет

$$R^*_{xx_{\text{ср}}}(\lambda)=\frac{1}{N}\sum_{i=1}^{N} R^i_{xx_{\text{ср}}}(\lambda), \qquad (7)$$

где N — количество реализаций в ансамбле.

Взаимная корреляционная функция, связывающая значение процесса $x_i(t)$ в момент времени t и значение процесса $y_i(t)$ в момент $t+\lambda$, имеет вид

$$R^{i,j}_{xy}(t,\lambda)=\frac{1}{T_3}\int\limits_{t-\frac{T_3}{2}}^{t+\frac{T_3}{2}} x^0_i(t)y^0_j(t+\lambda)\mathrm{d}t, \qquad (8)$$

где $y^0_j(t)=y_j(t)-M\left(y_j(t)\right)$, $M\left(y_j(t)\right)=\frac{1}{T}\int\limits_{t-\frac{T}{2}}^{t+\frac{T}{2}} y_j(t)\mathrm{d}t$ — математическое ожидание процесса $y_j(t)$, $y_j(t)$ — реализации исходных случайных процессов, $i\neq j$; T, T_3 — интервалы текущего сглаживания.

Средняя взаимная корреляционная функция

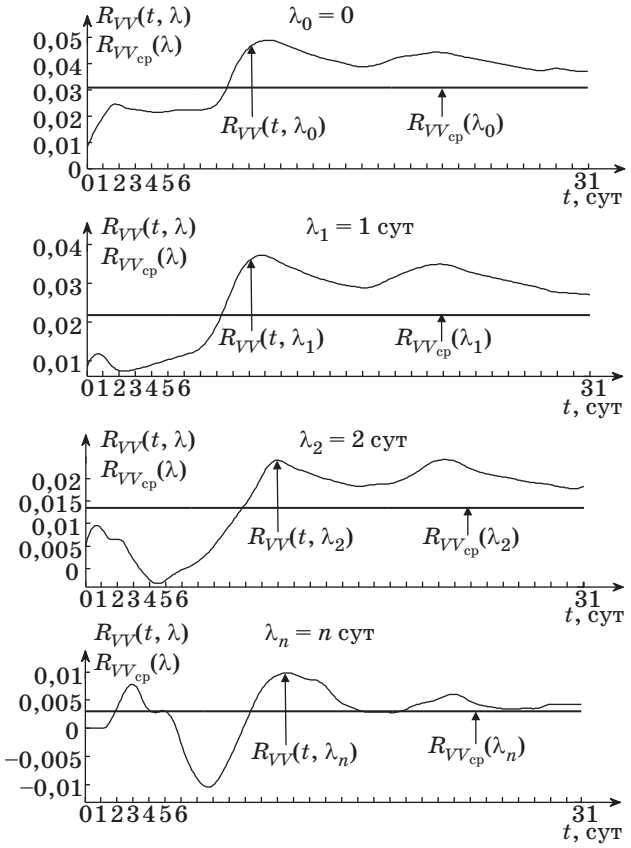
$$R^{i,j}_{xy_{\text{ср}}}(\lambda)=\frac{1}{T_2}\int\limits_0^{T_2} R^{i,j}_{xy}(t+\lambda)\mathrm{d}t, \qquad (9)$$

где T_2 — время усреднения аналогично (6).

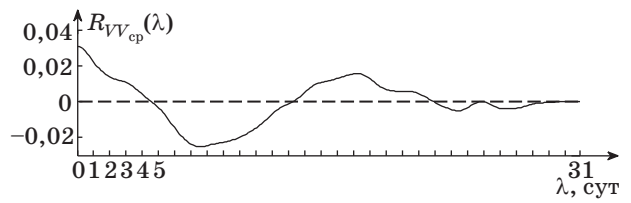
На рис. 5, 6 представлены примеры результатов обработки НСП по $V(t)$ на основе алгоритмов (4)–(7) и их дискретизации, приведенной в работе [6].

Обработка проводилась в среде MatLab при помощи приложения, аналогичного созданному в подсистеме САПР ЗСУПТ «Моделирование» [7].

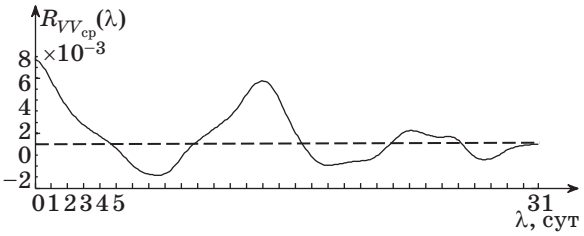
Составляющие $R_{TT}(\lambda_\mu)$, $R_{PP}(\lambda_\mu)$, $R_{WW}(\lambda_\mu)$ автокорреляционных функций температуры, давления, влажности и их средние значения $R_{TT_{ср}}(\lambda)$, $R_{PP_{ср}}(\lambda)$, $R_{WW_{ср}}(\lambda)$ вычисляются аналогично таким же характеристикам скорости ветра. На рис. 7 показан результат вычисления средней автокорреляционной функции $R_{TT_{ср}}(\lambda)$ температуры на основе алгоритмов (4)–(7).



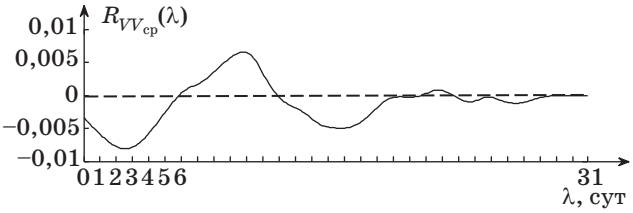
■ ***Рис. 5.*** Составляющие $R^i_{VV}(t, \lambda_\mu)$ автокорреляционной функции скорости ветра и их средние значения $R^i_{VV_{ср}}(\lambda_\mu)$



■ ***Рис. 6.*** Средняя автокорреляционная функция $R_{VV_{ср}}(\lambda)$ скорости ветра после приведения к эквивалентному стационарному процессу



■ ***Рис. 7.*** Средняя автокорреляционная функция $R_{TT_{ср}}(\lambda)$ температуры



■ ***Рис. 8.*** Средняя взаимная корреляционная функция $R_{VT_{ср}}(\lambda)$ между скоростью ветра и температурой

Результат вычисления средней взаимной корреляционной функции $R_{VT_{ср}}(\lambda)$ между скоростью ветра и температурой на основе алгоритмов (8), (9) представлен на рис. 8.

Алгоритмизация автоподстройки ПИД-регулятора

После расчета характеристик НСП по V, T^0, P, W составим алгоритм, реализующий зависимость (3).

Поскольку функция $K_и = F^0(V, T^0, P, W, S)$ неизвестна, то приближение к ней можно построить, исходя из следующего соотношения:

$$\tau_{потока} = \frac{S}{M(\tilde{V}(t))} \cong \tau_{ср} + \Delta\tau, \qquad (10)$$

где $M\left(\tilde{V}(t)\right) = M\left(V(t)\right) + M\left(\Delta\tilde{V}(t)/T^0, P, W\right)$, $M(V(t))$, $t_i < t < t_{i+1}$, $i = 1, ..., 12$ (месяцы) — математическое ожидание усредненной скорости потока по основной составляющей — скорости ветра $V(t)$ — с учетом колебаний по месяцам, соответствующее $\tau_{ср}$; $M\left(\Delta\tilde{V}(t)/T^0, P, W\right)$ — условное математическое ожидание дополнения к усредненной скорости потока, соответствующее $\Delta\tau$ — дополнению к $\tau_{ср}$ от влияния остальных метеорологических составляющих T^0, P, W, при этом t меняется в диапазоне $t_i < t < t_{i+1}$ (сутки), в предельном случае $\{t = 0, 1\}$ (полдень — 0, полночь — 1).

Оценка $M(V(t))$ производится по результатам обработки реализаций $V(t)$, оценка величины $M\left(\Delta\tilde{V}(t)/T^0, P, W\right)$ — по уравнению регрессии:

$$M\left(\Delta\tilde{V}(t)/T^0,P,W\right)=b_0+\sum_1^3 b_i\cdot M\left(X_i(t)\right),\quad (11)$$

$X_1 = T^0$, $X_2 = P$, $X_3 = W$; b_0, b_i — коэффициенты регрессии.

Определение коэффициентов b_0, b_i осуществляется по методу наименьших квадратов:

$$Q=\left[M\left(\Delta\tilde{V}(t)/x_i\right)-\left\{b_0+\sum_1^3 b_i\cdot M(x_i)\right\}\right]^2\to\min,$$

$$\text{тогда } \frac{\partial Q}{\partial b_i}=0,\ i=0,1,2,3;$$

$$-\frac{1}{2}\frac{\partial Q}{\partial b_0}=M\left(\Delta\tilde{V}(t)/x_i\right)-\left\{b_0+\sum_1^3 b_i\cdot M(x_i)\right\}=0;$$

$$-\frac{1}{2}\frac{\partial Q}{\partial b_i}=\left\{M\left(\Delta\tilde{V}(t)/x_i\right)-\left(b_0+\sum_1^3 b_i\cdot M(x_i)\right)\cdot M(x_m)\right\}=0,$$

$$i=1,2,3;\ m=1,2,3.\quad (12)$$

Из (12) получаем систему уравнений

$$M\left(\Delta\tilde{V}(t)/x_i\right)-b_0-b_1M(x_1)-b_2M(x_2)-b_3M(x_3)=0;$$

$$M\left(\Delta\tilde{V}(t)/x_i\right)-b_0M(x_1)-b_1\left\{M(x_1)\right\}^2-$$

$$-b_2M(x_2)M(x_1)-b_3M(x_3)M(x_1)=0;$$

$$M\left(\Delta\tilde{V}(t)/x_i\right)-b_0M(x_2)-b_1M(x_1)M(x_2)-$$

$$-b_2\left\{M(x_2)\right\}^2-b_3M(x_3)M(x_2)=0;$$

$$M\left(\Delta\tilde{V}(t)/x_i\right)-b_0M(x_3)-b_1M(x_1)M(x_3)-$$

$$-b_2M(x_2)M(x_2)-b_3\left\{M(x_3)\right\}^2=0.\quad (13)$$

Представим (13) в матричной форме

$$\mathbf{AB}=\mathbf{C},\quad (14)$$

где

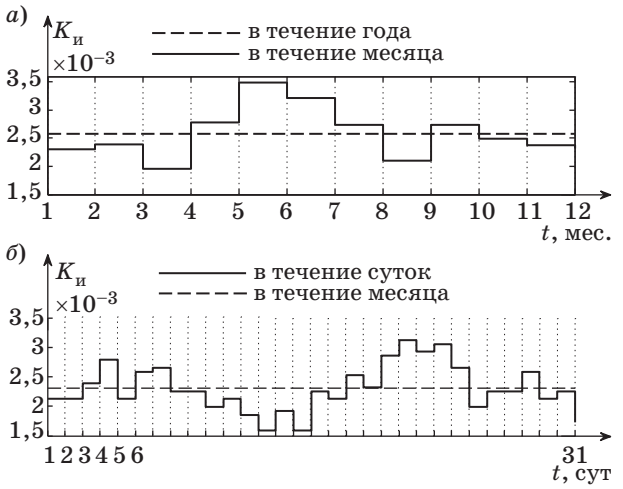
$$\mathbf{A}=\begin{bmatrix}1 & M(x_1) & M(x_2) & M(x_3)\\ M(x_1) & \{M(x_1)\}^2 & M(x_2)M(x_1) & M(x_3)M(x_1)\\ M(x_2) & M(x_1)M(x_2) & \{M(x_2)\}^2 & M(x_3)M(x_2)\\ M(x_3) & M(x_1)M(x_3) & M(x_2)M(x_3) & \{M(x_3)\}^2\end{bmatrix};$$

$$\mathbf{B}=\begin{bmatrix}b_0\\ b_1\\ b_2\\ b_3\end{bmatrix};\ \mathbf{C}=\begin{bmatrix}M\left(\Delta\tilde{V}(t)/x_i\right)\\ M\left(\Delta\tilde{V}(t)/x_i\right)M(x_1)\\ M\left(\Delta\tilde{V}(t)/x_i\right)M(x_2)\\ M\left(\Delta\tilde{V}(t)/x_i\right)M(x_3)\end{bmatrix}.\quad (15)$$

Элементы вектора **C** принимаются в виде

$$M\left(\Delta\tilde{V}(t)/x_i\right)\cong M\left(\bar{D}_{\text{ср}}\left(V(t)/x_1\right)\right),\ i=1,2,3,$$

где $M\left(\bar{D}_{\text{ср}}\left(V(t)/x_1\right)\right)$ — условное математическое ожидание средней дисперсии ветра $V(t)$; $\bar{D}_{\text{ср}}\left(V(t)/x_1\right)$ определяется по взаимно корреляционным функциям (8), (9).



■ ***Рис. 9.*** Закон управления автоподстройкой ПИД-регулятора с усреднением по месяцам (*а*) и суткам (*б*)

Проведем нормирование величин M, S в соотношении

$$K_{\text{и}}=\frac{a\left[M\left(V(t)\right)+M\left(\Delta\tilde{V}(t)/T^0,P,W\right)\right]}{S}$$

в виде

$$\bar{M}\left(V(t)\right)=\frac{M\left(V(t)\right)}{\max\left\{MV(t)\right\}};\ \bar{S}=\frac{S}{S_{\max}};$$

$$\bar{M}\left(\Delta\tilde{V}(t)/x_i\right)=\frac{M\left(\Delta\tilde{V}(t)/x_i\right)}{\max\left\{M\left(\Delta\tilde{V}(t)/x_i\right)\right\}}.\quad (16)$$

После нормирования (16) получим алгоритм управления автоподстройкой коэффициента $K_{\text{и}}$ в ПИД-регуляторе с учетом ввода в ЗСУПТ метеорологических данных:

$$K_{\text{и}}=\frac{a\left[\bar{M}\left(V(t)\right)+\bar{M}\left(\Delta\tilde{V}(t)/T^0,P,W\right)\right]}{\bar{S}}.\quad (17)$$

Закон управления автоподстройкой во времени для конкретных численных значений представлен на графиках (рис. 9, *а*, *б*).

Расчеты проводились при значениях параметров в (17), $a=1$, $S=1500$ м, $M\left(\tilde{V}(t)\right)$, $M\left(\Delta\tilde{V}(t)/x_i\right)$ вычислялись по алгоритмам (8), (9) и решению уравнений (13).

Численные значения $M\left(\tilde{V}(t)\right)$, $M\left(\Delta\tilde{V}(t)\right)$ вычислялись после приведения НСП V, T^0, P, W к эквивалентным стационарным процессам в соответствии с алгоритмами (4)–(9).

Заключение

Таким образом, предложенные в статье алгоритмы, а также сформированные законы управления автоподстройкой ПИД-регулятора позволяют осуществлять коррекцию управления при проектировании и производстве ЗСУПТ по многим критериям [8].

Литература

1. **Сольницев Р. И.** Построение замкнутой системы «Природа — Техногеника» // Информационные технологии в науке, образовании, телекоммуникации и бизнесе. IT + S&E`06: материалы XXXIII Междунар. конф., IV Междунар. конф. молодых ученых, Украина, Крым, Ялта-Гурзуф, 20–30 мая 2006 г. — Запорожье: Запорож. нац. ун-т, 2006. (Приложение к журналу «Открытое образование»). С. 404–408.
2. **Сольницев Р. И.** Вопросы построения замкнутой системы управления «Природа-Техногеника» // Изв. СПбГЭТУ «ЛЭТИ». 2009. № 7. С. 23–32.
3. **Сольницев Р. И., Коршунов Г. И.** Системы управления «Природа—Техногеника». — СПб.: Политехника, 2013. — 205 с.
4. **National** Hydro-meteorological Service National Center for Hydro-meteorological Forecasting (NCHMF). http://www.thoitietvietnam.gov.vn/web/en-US/62/19/58/map/Default.aspx (дата обращения: 17.01.2013).
5. **Пугачев В. С.** Введение в теорию вероятностей. — Наука. Глав. ред. физико-математической литературы, 1968. — 368 с.
6. **Сольницев Р. И.** Вычислительные машины в судовой гироскопии. — Л.: Судостроение, 1977. — 312 с.
7. **Сольницев Р. И., Тревгода М. А.** Программное обеспечение подсистемы САПР замкнутой системы управления «Природа-техногеника» // Информационно-управляющие системы. 2010. № 4. С. 34–38.
8. **Коршунов Г. И., Тисенко В. Н.** Управление процессами и принятие решений: учеб.-метод. пособие. — СПб.: СПбГПУ, 2010. — 231 с.

УДК 621.396.96

МОДЕЛИ СИГНАЛОВ В РАДИОПОЛЯРИМЕТРИИ

О. Д. Москалец,
канд. техн. наук, старший научный сотрудник
Санкт-Петербургский государственный университет аэрокосмического приборостроения

Предлагается векторная модель сигналов при исследовании поляризационных характеристик электромагнитных волн при их классическом описании. Введены поляризационные спектры векторных сигналов, где каждая бесконечно малая векторная монохроматическая компонента имеет свое, индивидуальное состояние поляризации. Установлено существование границ классического, приближенного, описания электромагнитных сигналов во временной и частотной областях. Показана необходимость развития физического аспекта теории сигналов.

Ключевые слова *— информация, сигнал, скалярная модель сигнала, векторная модель сигнала, состояние поляризации, поляризационный спектр, вектор Джонса, физический аспект.*

Введение

Проблемы извлечения, передачи и обработки информации являются центральными для многих областей науки и техники, таких как связь, автоматическое управление, радиолокация и радионавигация, оптика, радиофизика. В настоящее время к этим областям следует добавить радиочастотную идентификацию, методы которой находят все большее применение в различных сферах деятельности [1–3]. Успешное решение отмеченных проблем возможно при условии установки модели сигнала, позволяющей адекватно отображать информацию, переносимую сигналом.

Основным физическим носителем информации в названных областях науки и техники являются электромагнитные (ЭМ) волны, имеющие векторный характер, и для их полного описания необходимо кроме амплитуды, частоты и фазы указать поляризацию волны. Состояние поляризации является дополнительным информационным параметром, который не нашел достаточного отражения в теории информации и в теории сигналов (ТС). В этих областях моделью динамических сигналов является скалярная функция времени $s(t)$, которая может описывать колебания электрической и магнитной компонент ЭМ-поля, электрического тока или напряжения. Скалярной функции $s(t)$ соответствует ЭМ-волна с одной постоянной поляризацией как при передаче, так и при приеме, где ЭМ-излучение преобразуется в скалярный сигнал. При этом теряется та или иная информация, переносимая сигналом. Для получения максимального информационного содержания ЭМ-волны необходимо учитывать ее поляризационные свойства.

Скалярные модели динамических сигналов

Всякая обработка радиосигналов, в том числе и поляризационные преобразования, должна исходить из принятой модели сигнала. Основное требование, предъявляемое к модели сигнала, — это адекватное отображение информации, переносимой сигналом. Одним из требований модели сигнала является физическая содержательность. Поэтому построение моделей сигналов должно быть тесно связано с классом систем, порождающих исследуемые сигналы. Выдвижение модели сигнала исходит из двух непреложных постулатов: энергия сигнала и его протяженность во времени ограничены.

Адекватной моделью динамических сигналов является финитный нестационарный случайный процесс длительности T [4]. Случайный процесс $X_T(t)$, моделирующий сигнал, рассматривается как гармонизуемый, т. е. представимый в форме интеграла Фурье — Стилтьеса:

$$X_T(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(i\omega t)\mathrm{d}Z(\omega), \qquad (1)$$

где t — время; $Z(\omega)$ — случайная спектральная функция; ω — угловая временная частота.

Случайная спектральная функция $Z(\omega)$ в соотношении (1) дифференцируема почти наверное,

т. е. на всем множестве реализаций с вероятностью единица существует комплексная случайная спектральная функция [5]

$$S(\omega)=\frac{\mathrm{d}Z(\omega)}{\mathrm{d}\omega}, \qquad (2)$$

которая рассматривается как множество $\{{}^kS(\omega)\}$ реализаций комплексных спектральных функций. Это позволяет записать реализацию ${}^kx(t)$ нестационарного случайного процесса $X_T(t)$ в форме

$${}^kx(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}{}^kS(\omega)\exp(i\omega t)\mathrm{d}\omega, \qquad (3)$$

где

$${}^kS(\omega)=\int_{-\frac{T}{2}}^{\frac{T}{2}}{}^kx(t)\exp(i\omega t)\mathrm{d}t \qquad (4)$$

— комплексный спектр реализации ${}^kx(t)$.

В дальнейшем будут рассматриваться единственные реализации ЭМ-сигнала $S(\omega)={}^kS(\omega)$ в частотном пространстве и $s(t)={}^kx(t)$ как функция времени, которые наблюдаются в условиях реального физического эксперимента, и под функцией $s(t)$ понимаются колебания любой из скалярных компонент E_x, E_y, E_z электрического вектора ЭМ-поля:

$$\mathbf{E}(\mathbf{r},t)=\mathbf{i}E_x(\mathbf{r},t)+\mathbf{j}E_y(\mathbf{r},t)+\mathbf{k}E_z(\mathbf{r},t), \qquad (5)$$

где $\mathbf{r}=(x, y, z)$ — радиус-вектор; $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ — орты, связанные с осями x, y, z; E_x, E_y, E_z — проекции вектора $\mathbf{E}(\mathbf{r}, t)$ на оси x, y, z.

Соотношения (3) и (4) полностью согласуются с детерминистическими представлениями классической электродинамики, где вектор электрической компоненты $\mathbf{E}(\mathbf{r}, t)$ ЭМ-поля удовлетворяет волновому уравнению

$$\Delta\mathbf{E}-\frac{1}{\upsilon^2}\frac{\partial^2\mathbf{E}}{\partial t^2}=0, \qquad (6)$$

из которого следует, что функция $\mathbf{E}(\mathbf{r}, t)$ является дважды дифференцируемой по пространственным координатам и по времени, а потому непрерывна и ограничена. Это означает, что финитные компоненты в разложении (5) удовлетворяют условиям Дини и представимы в форме двойного интеграла Фурье:

$$s(t)=\frac{1}{2\pi}\upsilon p\int_{-\infty}^{\infty}\mathrm{d}\omega\int_{-\infty}^{\infty}s(t')\exp[i\omega(t-t')]\mathrm{d}t', \qquad (7)$$

где υp означает главное значение интеграла по переменной ω, что далее не оговаривается.

В форме (7) также представимы колебания электрического тока $i(t)$ или напряжения $u(t)$, которые даются известными соотношениями теории электричества:

$$i=\int_{\Delta S}\mathbf{j}_{\text{пр}}\mathrm{d}\mathbf{s},\ \ \mathbf{j}_{\text{пр}}=\sigma\mathbf{E},\ \ u_{ab}=\int_a^b\mathbf{E}\mathrm{d}\mathbf{l}, \qquad (8)$$

где ΔS — площадь поперечного сечения проводника; $\mathbf{j}_{\text{пр}}$ — вектор плотности тока проводимости; σ — удельная проводимость.

Поскольку в выражении (8) интегрирование выполнялось по пространственным координатам, то аналитические свойства функций $i(t)$ и $u(t)$ такие же, как у функции $\mathbf{E}(t)$. Поэтому не нужно постулировать, что колебания ЭМ-природы удовлетворяют условиям Дирихле.

Спектр финитной функции (4) в соответствии с интерполяционной теоремой Уиттекера представим в форме

$$S(\omega)=\sum_{-\infty}^{\infty}S\left(\frac{n\pi}{T}\right)\cdot\frac{\sin(\omega T-n\pi)}{(\omega T-n\pi)}. \qquad (9)$$

Эта теорема позволяет перейти от континуального представления спектра $S(\omega)$ к бесконечному счетному множеству

$$\{S(\omega_n)\}=\left\{S\left(\frac{n\pi}{T}\right)\right\}, \qquad (10)$$

которое содержит всю информацию о спектральном составе импульсного финитного сигнала и может быть полезно при рассмотрении многоканальной обработки сигналов.

Векторная модель динамического сигнала в форме электромагнитного поля

Описание состояния поляризации ЭМ-волны и ее поляризационных преобразований дают, исходя из плоской волны [6]; такая же волна предполагается при выдвижении векторной модели ЭМ-сигнала, далее эта волна предполагается однородной.

Однородная плоская ЭМ-волна, распространяющаяся вдоль оси z в декартовой системе координат x, y, z, может быть представлена в виде [6]

$$\mathbf{E}(x,y,z,t)=(\mathbf{i}\dot{E}_x+\mathbf{j}\dot{E}_y)\exp[i(\omega t-kz)]=$$
$$=\begin{bmatrix}\dot{E}_x\\ \dot{E}_y\end{bmatrix}\exp[i(\omega t-kz)], \qquad (11)$$

где $\dot{E}_x, \dot{E}_y$ — комплексные амплитуды горизонтальной и вертикальной компоненты соответственно; $k=\omega/c$ — волновое число; c — скорость света.

Матрица-столбец в последней формуле цепи равенств (11) является вектором Джонса однородной монохроматической плоской волны, который определяет ее состояние поляризации. Выдвижение векторной модели сигнала опирается на то обстоятельство, что электромагнитная монохро-

матическая волна всегда поляризована — эллиптически, циркулярно или линейно.

В линейной среде спектральные компоненты сигнала распространяются независимо друг от друга, и поведение скалярной волны дается суперпозицией гармонических волн бесконечно малой амплитуды:

$$s(z,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S(\omega)\exp[i(\omega t-kz)]\mathrm{d}\omega. \qquad (12)$$

Соотношение (12) пригодно для описания частного случая ЭМ-волны при ее вертикальной или горизонтальной поляризации. Общий случай состояния поляризации ЭМ-поля требует введения векторной модели динамического сигнала [7].

Векторная модель сигнала [7] исходит из того, что в форме (12) представима и горизонтальная, и вертикальная компоненты у плоского ЭМ-поля. Тогда векторный сигнал запишется в виде

$$\mathbf{E}(t,z)=\mathbf{i}\frac{1}{2\pi}\int_{-\infty}^{\infty}S_x(\omega)\exp[i(\omega t-kz)]\mathrm{d}\omega+$$

$$+\mathbf{j}\frac{1}{2\pi}\int_{-\infty}^{\infty}S_y(\omega)\exp[i(\omega t-kz)]\mathrm{d}\omega=$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left\{\left[\mathbf{i}S_x(\omega)+\mathbf{j}S_y(\omega)\right]\exp[i(\omega t-kz)]\right\}\mathrm{d}\omega. \qquad (13)$$

Соотношению (13) соответствует представление в форме

$$\mathbf{E}(t,z)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\begin{bmatrix}S_x(\omega)\\S_y(\omega)\end{bmatrix}\exp[i(\omega t-kz)]\mathrm{d}\omega, \qquad (14)$$

где

$$\left|S_x(\omega)\right|^2+\left|S_y(\omega)\right|^2=\left|\Phi(\omega)\right|^2. \qquad (15)$$

Матрица-столбец в выражении (14) определяет вектор Джонса поляризационных спектров [7]

$$\mathbf{J}_s=\begin{bmatrix}S_x(\omega)\\S_y(\omega)\end{bmatrix}, \qquad (16)$$

который, по-видимому, введен в работе [8] и в дальнейшем был подтвержден в публикациях [9, 10].

Подобно тому, как скалярное соотношение (12) является суперпозицией скалярных колебаний, выражения (13) и (14) представляют собой суперпозицию векторных колебаний. В этих выражениях полагается, что каждая пара бесконечно малых спектральных волновых компонент с угловой частотой ω

$$S_x(\omega)\exp[i(\omega t-kz)]\mathrm{d}\omega,\ \ S_y(\omega)\exp[i(\omega t-kz)]\mathrm{d}\omega \qquad (17)$$

имеет свое индивидуальное состояние поляризации (эллиптическое, циркулярное или линейное), и в совокупности эти компоненты составляют векторный сигнал (14) с теми или иными поляризационными свойствами — от полной поляризации до полного ее отсутствия.

Соотношения (13) и (17) позволяют записать следующее выражение:

$$\mathrm{d}\mathbf{E}(t,z)=\left\{\left[\mathbf{i}S_x(\omega)+\mathbf{j}S_y(\omega)\right]\exp[i(\omega t-kz)]\right\}\mathrm{d}\omega=$$

$$=\begin{bmatrix}S_x(\omega)\\S_y(\omega)\end{bmatrix}\exp[i(\omega t-kz)]\mathrm{d}\omega, \qquad (18)$$

которое аналогично соотношению (11), но для бесконечно малых величин.

Вектор Джонса (16) описывает состояние поляризации исходного импульсного векторного ЭМ-сигнала, поляризационные характеристики которого преобразовываются прибором, изменяющим состояние поляризации. Подобно тому, как формулы (3) и (4) представляют реализации соответствующих случайных функций, конкретный вектор Джонса (16) также представляет одну из реализаций состояния поляризации случайного векторного ЭМ-сигнала [8]. Иными словами, для стохастического описания векторного сигнала необходимо иметь в виду вероятностный характер и скалярного передаваемого сигнала $s(t)$, и вектора (16) одновременно [8].

Математические модели и физическая сущность динамических сигналов

Для теории информации, теории сигналов и теории линейных систем большое значение имеет существование интегралов

$$I=\int_{-\infty}^{\infty}\frac{\left|\log G(\omega)\right|}{1+\omega^2}\mathrm{d}\omega;\ \ I=\int_{-\infty}^{\infty}\frac{\log\left|K(\omega)\right|}{1+\omega^2}\mathrm{d}\omega, \qquad (19)$$

где $G(\omega)$ — энергетический спектр стационарного случайного процесса; ω — безразмерная угловая частота; $K(\omega)$ — передаточная функция линейной стационарной системы.

Для сигналов с финитным спектром первый интеграл в формулах (19) расходится, и это означает детерминированность процесса, откуда делается вывод, что сигналы с финитным спектром (по определению А. Н. Колмогорова, сингулярные процессы [11]) не могут быть носителями информации [4]. Второй интеграл в формулах (19) известен как критерий Винера — Пэли реализации линейной физической системы. Если этот интеграл расходится, то физическая реализация линейной системы и формирование системой сигнала с финитным спектром невозможны [4].

Классическая теория не обратила внимания на следующий физический факт: спектр тормозного рентгеновского излучения как случайного

процесса — финитный [14], и это обстоятельство имеет фундаментальное значение для физики, так как позволяет экспериментально установить постоянную Планка, которая является одной из мировых констант [14]. Далее, верхняя граничная частота упругих колебаний в кристаллах определяется первой зоной Бриллюэна [15], и кристалл, по понятиям квантовой акустики, является фильтром нижних частот с обозначенной верхней граничной частотой полосы пропускания. Обозначенная ситуация требует специального рассмотрения, поскольку сигнал, согласно определению теории информации, является материальным, физическим носителем информации и как физический объект требует рассмотрения с физической точки зрения [16, 17] с привлечением квантовых представлений, так как квантовая физика дает точное описание физических явлений, а классические представления являются приближенными.

Преобразование Фурье финитной функции (9), согласно теореме Винера — Пэли, описывается целой функцией экспоненциального типа степени $T/2$. Такая функция имеет бесконечную протяженность по оси угловых частот ω и обращается в ноль лишь в точках, которые являются корнями этой функции. При этом вопрос о физическом смысле частот $\omega \to \infty$ ТС и теория информации не ставят.

Энергия сигнала W, выраженная в форме теоремы Парсеваля, должна иметь конечное значение:

$$W = \int_{-\infty}^{\infty} s^2(t)\mathrm{d}t = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\omega)\mathrm{d}\omega = \frac{1}{\pi}\int_{0}^{\infty} G(\omega)\mathrm{d}\omega, \quad (20)$$

где $G(\omega) = |S(\omega)|^2$ — энергетический спектр сигнала $s(t)$.

Последний интеграл в цепи равенств (20) можно представить в форме

$$\int_{0}^{\infty} G(\omega)\mathrm{d}\omega = \int_{0}^{\omega_N} G(\omega)\mathrm{d}\omega + \int_{\omega_N}^{\infty} G(\omega)\mathrm{d}\omega = W_N + \Delta W. \quad (21)$$

В силу сходимости первого интеграла в выражении (21)

$$\Delta W = \int_{\omega_N}^{\infty} G(\omega)\mathrm{d}\omega \to 0 \text{ при } \omega_N \to \infty, \quad (22)$$

тогда при достаточно больших значениях ω_N действие

$$\Delta W \cdot T_0 \le \hbar, \quad (23)$$

где T_0 — некоторая временная величина, связанная с данным сигналом, например, $T_0 = T$; $\hbar$ — постоянная Планка.

Неравенство (23) указывает на то, что при достаточно больших величинах ω_N классическое описание спектра $S(\omega)$ финитного сигнала в форме целой функции экспоненциального типа становится неприемлемым, что следует из общефизического принципа, устанавливающего границы классического описания [14]. Иными словами, классическое представление реализации финитного сигнала в форме (3) возможно лишь в конечной полосе частот.

С другой стороны, если сигнал имеет финитный спектр с носителем $\mathrm{supp}S(\omega) = [-\Omega, \Omega]$, т. е. представим в виде

$$s(t) = \frac{1}{2\pi}\int_{-\Omega}^{\Omega} S(\omega)\exp(i\omega t)\mathrm{d}\omega, \quad (24)$$

то, согласно теореме Винера — Пэли, сигнал $s(t)$ описывается целой функцией экспоненциального типа степени Ω.

Энергию сигнала (24) можно представить в форме

$$W = \int_{-\infty}^{0} s^2(t)\mathrm{d}t + \int_{0}^{\infty} s^2(t)\mathrm{d}t = W_1 + W_2, \quad (25)$$

в свою очередь

$$W_1 = \int_{-t_N}^{0} s^2(t)\mathrm{d}t + \int_{-\infty}^{-t_M} s^2(t)\mathrm{d}t = W_{1M} + \Delta W_1; \quad (26)$$

$$\int_{0_N}^{t_M} s^2(t)\mathrm{d}t + \int_{t_M}^{\infty} s^2(t)\mathrm{d}t = W_{2M} + \Delta W_2. \quad (27)$$

В силу сходимости интегралов в выражении (25)

$$\Delta W_1 + \Delta W_2 = \int_{-\infty}^{t_N} s^2(t)\mathrm{d}t + \int_{t_M}^{\infty} s^2(t)\mathrm{d}t \to 0$$

$$\text{при } t_N \to -\infty,\ t_M \to \infty. \quad (28)$$

В этом случае при достаточно больших величинах $|t_N|$ и t_M действие

$$(\Delta W_1 + \Delta W_2)\cdot T_0 \le \hbar. \quad (29)$$

Неравенство (29) указывает на то, что при достаточно малых уровнях сигналов их описание в классической форме становится также неприемлемым. Иначе говоря, классическое описание сигнала с финитным спектром возможно на ограниченном промежутке времени.

Неравенства (15) и (21) являются математическим обоснованием известных воззрений о том, что классическая электродинамика — это электродинамика сравнительно низких частот [12] и сильных полей [13]. Эти воззрения и неравенства (23) и (29) позволяют заключить, что при классическом, приближенном, описании динамических сигналов сигналы с финитным спектром должны мыслиться как финитные, и это не столь-

ко техническое, сколько физическое обстоятельство. По этой причине счетное множество (11) должно быть ограниченным.

Квантовая электродинамика дает выражение энергии ЭМ-поля в одномерном информационном канале в виде суммы

$$W = \sum_i N_i \hbar \omega_i, \qquad (30)$$

где N_i — квантовые числа; ω_i — частота фотона.

Формула (30), по существу, представляет распределение энергии ЭМ-излучения по частотам в форме дискретной функции. Очевидно, счетное множество $\{\omega_i\}$ должно быть ограниченным, в противном случае энергия сигнала станет бесконечной.

Вероятностный смысл квантовой физики и процесс формирования сигнала при его точном, квантовом представлении [18] дают основание утверждать, что информационные свойства ЭМ-сигнала, т. е. его стохастичность, определяются не аналитическими свойствами спектральных функций сигнала, а квантовой природой ЭМ-излучения и, соответственно, формирования сигнала.

Заключение

Наряду с энергией и веществом информация стала важнейшим понятием для человечества, и сигнал, как материальный, физический, носитель информации, требует тщательного и всестороннего изучения — необходимо подробное исследование существующих моделей сигналов и выдвижение новых. При этом математические модели сигналов должны сообразовываться с физическим содержанием сигнала.

Исследования по ТС ведутся в двух основных направлениях: чисто математическое обоснование технических применений и выяснение физических основ. Первое направление, где сигналы мыслятся как математические объекты, составляет главную часть ТС. Оно развито значительно больше второго, и ТС рассматривается в качестве прикладного раздела функционального анализа, как «математика для радио». Попытки разработать второе направление сводились, по существу, к дальнейшему развитию первого направления. В результате физическая проблематика ТС осталась в стороне, при том, что под сигналом понимаются реальные физические процессы, происходящие в реальных физических системах.

В данной работе предпринята попытка, с одной стороны, развить математический аспект теории сигналов, с другой стороны, обратить внимание на физическую сторону вопроса. Результатом исследований явилась векторная модель сигнала, предложенная в рамках классических представлений. Здесь же было показано, что аналитические свойства существующих моделей ЭМ-сигналов вытекают из основных соотношений классической электродинамики. В то же время хорошо устоявшиеся классические представления требуют критической оценки.

Рассмотрение сигнала как физического объекта показало приближенный характер основных положений ТС, и это естественно, так как классическая теория дает приближенное описание физических явлений, а их точное описание дается в рамках квантовых представлений. Установленные границы классического описания сигнала во временной и частотной областях можно считать представленным здесь вкладом в развитие физического аспекта ТС. Эта тема отражена в целом ряде публикаций автора [16–19].

Развитие физического аспекта ТС становится все более актуальным в связи с широкими исследованиями в области фемтосекундных импульсов, которые являются квантовыми системами. Неэффективность применения классического описания ультракоротких (фемтосекундных) световых импульсов отмечена в работах [20, 21], и этот вопрос требует дальнейшего развития в рамках физических исследований.

Работа выполнена по государственному контракту № 14.527.12.0019; шифр лота 2011-2.7-527-025; шифр заявки 2011-2.7.-527-025-002.

Литература

1. **Койгеров А. С., Забузов С. А., Дмитриев В. Ф.** Исследование корреляционного метода для решения задач антиколлизии для систем радиочастотной идентификации на ПАВ // Информационно-управляющие системы. 2009. № 5. С. 48–55.
2. **Марковский С. Г., Марковская Н. В.** Разрешение конфликтов в системах радиочастотной идентификации с использованием идентификаторов меток и процедуры последовательной компенсации конфликтных сигналов // Информационно-управляющие системы. 2012. № 2. С. 48–55.
3. **Марковский С. Г., Марковская Н. В.** Расчет средней задержки алгоритма решения конфликтов в системах радиочастотной идентификации // Информационно-управляющие системы. 2012. № 4. С. 84–92.
4. **Железнов Н. А.** О принципиальных вопросах теории сигналов и задачах ее дальнейшего развития

на основе новой стохастической модели // Радиотехника. 1957. Т. 12. № 11. С. 3–12.
5. **Железнов Н. А.** Некоторые вопросы спектрально-корреляционной теории нестационарных сигналов // Радиотехника и электроника. 1959. Т. 4. № 3. С. 359–373.
6. **Джерард А., Бёрч Дж. М.** Введение в матричную оптику: пер. с англ. — М.: Мир, 1978. — 341 с.
7. **Москалец О. Д.** Модель сигнала при обработке векторных стохастических полей // Всесоюз. конф. по статистическим методам обработки данных дистанционного зондирования окружающей среды. Рига, 1986. С. 54.
8. **Москалец О. Д.** Учет поляризационных характеристик антенн при спектральных измерениях в радиоастрономии // Антенные измерения: IV Всесоюз. конф. «Метрологическое обеспечение антенных измерений» (ВКАИ-4). Ереван, 1987. С. 45–47.
9. **Козлов А. И., Логвин А. И., Сарычев В. А.** Поляризация радиоволн. Поляризационная структура радиолокационных сигналов. — М.: Радиотехника, 2005. — 704 с.
10. **Слетков В. Л.** Аналитическое представление поляризованных сигналов // Изв. вузов. Радиоэлектроника. 2006. Т. 49. № 3. С. 17–23.
11. **Колмогоров А. Н.** Теория передачи информации. — М.: Изд-во АН СССР, 1956. — 33 с.
12. **Гольдштейн Л. Д., Зернов Н. В.** Электромагнитные поля и волны. Изд. 2-е, перераб. и доп. — М.: Сов. радио, 1971. — 664 с.
13. **Берестецкий В. Б., Лифшиц Е. М., Питаевский Л. П.** Квантовая электродинамика. Изд. второе, перераб. — М.: Наука, 1980. — 704 с.
14. **Вихман Э.** Квантовая физика. 3-е изд., испр. / пер. с англ. — М.: Наука, 1986. — 292 с.
15. **Блейкмор Дж.** Физика твердого тела. — М.: Мир, 1988. — 608 с.
16. **Москалец О. Д.** Модель динамического сигнала в теории информации и квантовая физика // X симп. по проблеме избыточности в информационных системах / ЛИАП. Л., 1989. Ч. I. С. 164–167.
17. **Москалец О. Д.** Методы квантовой физики в теории сигналов // Proc. Latvian Signal Processing Int. Conf. Riga, 1990. P. 42–46.
18. **Москалец О. Д.** Электромагнитные сигналы в квантовой электронике: квантовое описание и классическое приближение // Изв. вузов. Физика. 2001. Т. 44. № 10. С. 6–12.
19. **Москалец О. Д.** Фемтосекундные импульсы: классическое и квантовое описание // Лазеры, измерения, информация 2012: сб. докл. 22-й Междунар. конф. Т. 2. СПб.: Изд-во Политехнического ун-та, 2012. С. 71–81.
20. **Беленов Э. М., Назаркин А. В., Прокопович И. П.** Динамика мощного фемтосекундного импульса // Письма в ЖЭТФ. 1992. Е. 55. Вып. 4. С. 223–227.
21. **Shvartsburg A. B.** Time-domain optics of ultrafast waveform. — Oxford: Caledonia Press, 1996. — 208 p.

УДК 004.9

СИСТЕМА МОДЕЛИРОВАНИЯ СЕТЕВЫХ ПОМЕХ МУЛЬТИМЕДИЙНЫХ ПОТОКОВ

А. А. Рогов,
доктор техн. наук, профессор
А. Л. Забровский,
аспирант, ведущий программист
Петрозаводский государственный университет

Представлена система моделирования сетевых помех, влияющих на качество мультимедийных потоков, передаваемых в IP-сети. Данная система предназначена для исследования и тестирования новых мультимедийных сервисов и систем, а также проверки создаваемых критериев оценки качества мультимедийных потоков. Показано, что для определения результирующего качества мультимедийных потоков, передаваемых в реальном режиме времени, можно использовать параметры, полученные от плееров удаленных пользователей. Для оценки качества использовались время начала воспроизведения, минимальное количество кадров в секунду, максимальный скачок потери кадров и минимальный размер буфера в секундах, который был зафиксирован в течение всего воспроизведения мультимедийного потока.

Ключевые слова *— система, моделирование, сетевые помехи, эмулятор, WANem, сеть, качество, мультимедийный поток.*

Введение

Мультимедийные потоки в режиме реального времени сегодня активно транслируются через Интернет. Потоки передаются как между двумя точками сети, так и в режиме многоточечной связи. Передача мультимедийных потоков применяется в образовании и бизнесе. Видеотрансляции конференций, лекций и других мероприятий становятся обыденным делом [1].

Для создания он-лайн трансляции необходимо иметь три основных элемента системы: видеокодер, медиа-сервер и приложение плеер на стороне клиента. В свою очередь, во время передачи мультимедийных потоков через сеть Интернет, а именно при взаимодействии медиа-сервера с клиентами, на качество мультимедийных потоков оказывают влияние различные сетевые помехи, поэтому качество изображения и звука на стороне клиента в некоторых случаях не достаточно хорошее. Если для передачи мультимедийного трафика на транспортном уровне сетевой модели OSI используется протокол UDP, то ухудшение видео на стороне клиента может выглядеть в виде рассыпания изображения на мелкие квадраты по причине потери отдельных видеокадров. При использовании протокола транспортного уровня TCP ухудшения, как правило, выражены в задержках и замирании изображения. В случае большой потери пакетов воспроизведение видео перестанет проигрываться при использовании обоих протоколов.

Задача повышения и анализа качества передаваемого видеопотока является весьма актуальной. Для ее решения требуется изучить влияние различных сетевых параметров на качество передаваемых мультимедийных потоков. Особенно востребованы подобные исследования при создании и тестировании новых мультимедийных сервисов и систем. Данные о влиянии сетевых помех на качество видеопотока могут быть получены с помощью пассивного эксперимента, т. е. путем наблюдения за реальными потоками. Более перспективным выглядит проведение активного эксперимента, т. е. моделирование помех искусственным образом с помощью программного обеспечения (ПО) или специального оборудования.

Моделирование сетевых помех

На сегодняшний день существует оборудование и ПО для эмулирования сетевых помех, например, устройство Linktropy 5500 WAN Emulator или ПО WANem (Wide Area Network Emula-

tor). Оборудование Linktropy 5500 WAN Emulator, предлагаемое компанией Apposite Technologies, позволяет эмулировать ширину сетевого канала, задержку пакетов, джиттер, потерю пакетов, перегрузку сетевого канала в лабораторных условиях [2]. Одно из основных предназначений устройства — это тестирование разрабатываемых клиент-серверных приложений, мониторинг трафика. Например, Linktropy 5500 WAN Emulator может эмулировать сетевую задержку от 1 мс до 10 с. Настройка и управление устройством осуществляются через веб-интерфейс.

Программное обеспечение WANem тоже позволяет эмулировать различные характеристики сети [3]. Преимущество WANem перед Linktropy 5500 WAN Emulator заключается в том, что это ПО распространяемое бесплатно. В качестве недостатков можно отметить необходимость использования дополнительного компьютера для установки WANem, а также меньшую функциональность, например, отсутствие интерфейса мониторинга передаваемого трафика. Представленные сетевые эмуляторы не содержат инструментов для оценки параметров состояния плеера и качества мультимедийных потоков на стороне клиента.

Созданная система имитационного моделирования мультимедийных потоков с сетевыми помехами позволяет не только эмулировать различные типы мультимедийных потоков, но и оценивать их качество. Предлагаемая система включает в себя инструменты оценки результатов экспериментов.

Система моделирования сетевых помех (рис. 1) включает следующие элементы:

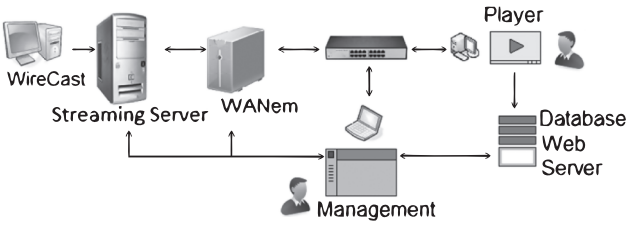
— веб-интерфейс управления (Management);

— базу данных Mysql (Database) и веб-сервер Apache (Web Server);

— медиа-сервер и видеокодер (Streaming Server и WireCast соответственно);

— сетевой эмулятор (WANem);

— Flash-медиа-плеер и компьютер клиента (Player).



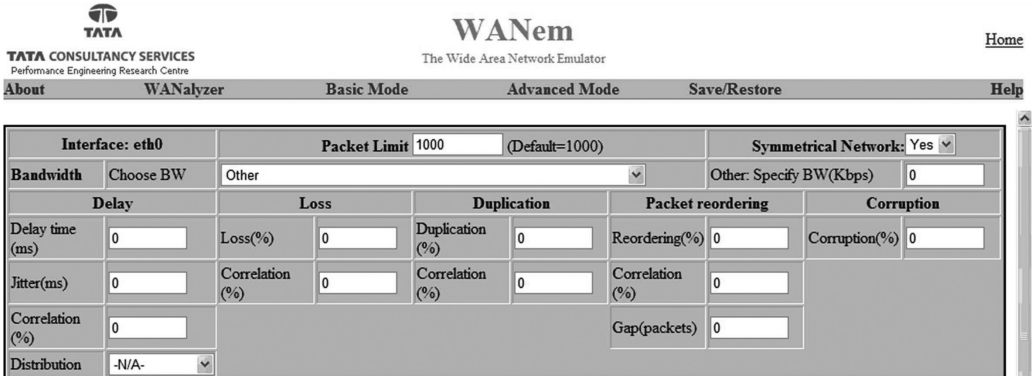
■ ***Рис. 1.*** Система моделирования сетевых помех

Для установки и использования системы достаточно одного рабочего стола. Размеры стола напрямую зависят от размеров используемых для экспериментов компьютеров.

Веб-интерфейс управления

Веб-интерфейс управления — это набор PHP-скриптов, хранящихся на веб-сервере. Функционал данного элемента системы обеспечивает просмотр списка проведенных экспериментов в веб-браузере с возможностью получения детальной информации о каждом проделанном эксперименте. Информация по каждому эксперименту выводится в виде интерактивных графиков, отражающих изменения различных параметров состояния воспроизведения плеера (рис. 2). Также в этот элемент входит веб-интерфейс настройки и запуска сетевого эмулятора WANem (рис. 3).

■ ***Рис. 2.*** Размер буфера Flash-плеера в секундах



■ ***Рис. 3.*** Веб-интерфейс настройки и запуска сетевого эмулятора WANem

Данный элемент системы позволяет добавлять описания эксперимента в базу данных. При этом существует разработанное стандартное описание экспериментов по умолчанию, в котором необходимо только указать значения эмулируемых сетевых параметров будущего эксперимента.

Для запуска эксперимента необходимо выбрать и загрузить через веб-интерфейс управления HTML-страницу со встроенным Flash-плеером. Flash-плеер запускается на компьютере клиента. Сразу после загрузки данной страницы интернет-браузером Flash-медиа-плеер начинает автоматически отсылать параметры состояния воспроизведения мультимедийного потока в удаленную базу данных Mysql.

База данных Mysql

База данных системы моделирования сетевых помех состоит из двух таблиц — *experiments* и *clients_parameters*. Таблица *experiments* включает информацию об экспериментах и содержит три столбца: *id_experiment* (идентификатор эксперимента), *experiment_name* (название эксперимента) и *description* (описание эксперимента).

Таблица *clients_parameters* содержит значения параметров воспроизведения мультимедийных потоков, полученные от Flash-медиа-плеера, например, количество потерянных видеокадров, текущий размер буфера плеера и др. Таблицы *experiments* и *clients_parameters* связаны между собой через *id_experiment* (идентификатор эксперимента).

Медиа-сервер и видеокодер

Компьютерная программа WireCast [4] выступает в роли видеокодера и предназначена для передачи на сервер мультимедийных потоков с разными битовыми скоростями в реальном режиме времени. Видеокодер WireCast установлен и работает на отдельном компьютере.

На медиа-сервере применяется ПО Flash Media Streaming Server. Для передачи мультимедийных потоков в режиме реального времени используется протокол RTMP (Real Time Messaging Protocol). Разработанная система моделирования сетевых помех позволяет задействовать разные медиа-серверы, например Wowza Media Server. Таким образом, возможно исследовать зависимость качества принимаемых мультимедийных потоков не только от сетевых ухудшений, но и от используемого в экспериментах медиа-сервера.

Сетевой эмулятор WANem

Сетевой эмулятор WANem установлен и настроен на сервере, который расположен между медиа-сервером и клиентом. Для того чтобы сетевой мультимедийный трафик передавался не напрямую от сервера к клиенту, а через WANem, необходимо прописать соответствующие маршруты в таблицах маршрутизации медиа-сервера и клиента. В предлагаемой системе используется расширенный режим настроек эмуляции (Advanced Mode) ПО WANem. В этом режиме доступно эмулирование различных характеристик сети, а также создание своих собственных правил.

Рассмотрим реальный пример того, как осуществляются настройка сетевого трафика и перенаправление его от медиа-сервера к клиенту и наоборот через сетевой эмулятор WANem (рис. 4). Ниже представлены IP-адреса медиа-сервера, сетевого эмулятора WANem и компьютера клиента:

— медиа-сервер (Flash Media Streaming Server): 10.33.17.243;

— сервер с сетевым эмулятором WANem: 10.33.17.81;

— клиентский компьютер с Flash-медиа-плеером: 10.33.17.242.

Сервер, на котором установлен сетевой эмулятор WANem, в представленной системе имеет только один сетевой интерфейс. Все три компьютера находятся в одной локальной сети и имеют IP-адреса из одной подсети.

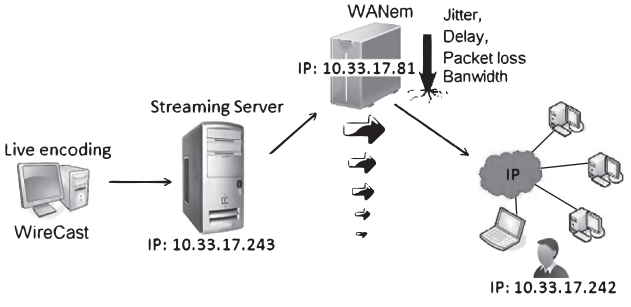
Для того чтобы пакеты с клиентского компьютера до сервера проходили через сетевой эмулятор WANem, а не напрямую, необходимо прописать маршрут в таблице маршрутизации клиентского компьютера. На компьютере с операционной системой Windows нужно открыть консоль (командная строка) и выполнить команду:

route add 10.33.17.243 mask 255.255.255.255 10.33.17.81

Для того чтобы пакеты с сервера до клиентского компьютера проходили через сетевой эмулятор WANem, на сервере необходимо выполнить следующую команду:

route add 10.33.17.242 mask 255.255.255.255 10.33.17.81

В данной схеме Flash-медиа-плеер подключается к серверу, получает и воспроизводит мультимедийный поток в реальном режиме времени. Таким образом, у экспериментатора есть возмож-



■ ***Рис. 4.*** Перенаправление сетевого трафика

ность эмулировать различные сетевые помехи с помощью WANem и влиять на качество передаваемых мультимедийных потоков.

Flash-медиа-плеер и компьютер клиента

Flash-медиа-плеер — это программный видеоплеер, который используется для воспроизведения мультимедийных потоков. Компьютер, на котором запускается Flash-медиа-плеер, обозначается как компьютер клиента и предназначается для визуального мониторинга качества воспроизводимых мультимедийных потоков. Параллельно с визуальным наблюдением экспериментатором качества воспроизводимого мультимедийного потока Flash-медиа-плеер отправляет свои собственные параметры состояния воспроизведения мультимедийного потока в удаленную базу данных. Для созданной системы в качестве плеера используется Strobe Media Playback [5] — свободно распространяемый Flash-медиа-плеер с открытым исходным кодом. Для экспериментов плеер был запрограммирован таким образом, чтобы перед началом воспроизведения мультимедийного потока он накапливал буфер данных, равный 10 с, и затем начинал отображение видеопотока.

В данной системе реализовано два варианта отправки плеером параметров в базу данных Mysql. Первый способ — посредством самого плеера, запрограммированного специально для этих целей, и второй способ — с помощью ПО, написанного на Java Script, которое в данном случае выступает в роли посредника меду плеером и базой данных. Интервал отправки параметров в базу данных считается в секундах и может изменяться экспериментатором. В качестве интернет-браузера используется Internet Explorer 9.0.

Подготовка, проведение и результаты эксперимента

Последовательность проведения эксперимента состоит из следующих шагов: в первую очередь, в веб-интерфейсе управления создается эксперимент и его описание, которые сохраняются в таблице *experiments* базы данных Mysql. Описание эксперимента включает в себя информацию об эмулируемых сетевых параметрах, характеристиках передаваемого потока и др.

Далее задается номер предстоящего эксперимента. Все операции взаимодействия с базой данных осуществляются через веб-интерфейс администратора.

На следующем шаге происходит настройка видеокодера и сетевого эмулятора WANem. В настройках видеокодера устанавливается битовая скорость транслируемого мультимедийного потока, например 1500 Кбит/с, а также остальные параметры кодирования потока (рис. 5). В качестве источника сигнала на входе видеокодера используется образовательное видео, созданное для экспериментов. Продолжительность видео составляет 182 с.

Через веб-интерфейс сетевого эмулятора WANem задаются значения сетевых параметров, которые будут выступать в роли сетевых помех.

После этого необходимо запустить трансляцию мультимедийного потока с видеокодера на сервер и загрузить HTML-страницу со встроенным Flash-плеером. Время проведения эксперимента чуть больше или равно продолжительности видеофайла.

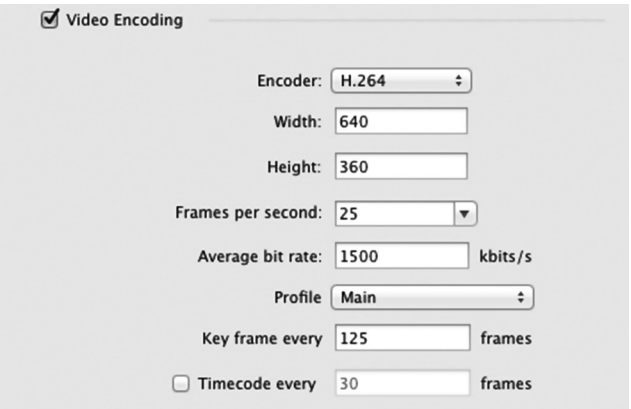
По окончании проведения эксперимента с помощью веб-интерфейса управления можно посмотреть графики изменения параметров воспроизведения мультимедийных потоков. Предусмотрена функция получения сводных оценок проведенных экспериментов при выборе и загрузке веб-страницы Results (результаты).

Определение величины битовых потоков, удовлетворяющих конечных пользователей, основано на анализе параметров, полученных от плееров удаленных пользователей. Были выбраны следующие четыре характеристики.

T_{start} — время начала воспроизведения — время, при котором первое значение количества воспроизводимых видеокадров в секунду больше 24.

$F_{\min}$ — минимальное количество кадров в секунду, которое было зафиксировано в течение воспроизведения мультимедиа-потока. Стандартное значение FPS = 25 кадрам в секунду для оригинального потока, но в некоторых случаях оно может уменьшаться; например, если процессор компьютера загружен и не успевает отображать все кадры, тогда происходит потеря кадров.

F_{drop} — максимальный скачок потери кадров, который был зафиксирован в течение воспроизведения мультимедийного потока.



■ ***Рис. 5.*** Параметры видеокодера

$B_{\min}$ — минимальный размер буфера в секундах, который был зафиксирован в течение всего воспроизведения мультимедийного потока.

Была поставлена задача: можно ли по этим четырем характеристикам определить качество мультимедийного потока.

Разработанная система применялась для планирования и проведения ряда экспериментов. На основе проведенных экспериментов были сформированы две выборки: обучающая и контрольная. Обучающая выборка включала данные по 129 экспериментам, а контрольная — по 66. Все эксперименты проводились для пяти потоков с разными битовыми скоростями.

Подбор вида функции основывался на физических соображениях об исследуемом процессе. На основе обучающей выборки было обнаружено, что дискриминантная функция вида

$$y = \frac{F_{\min} \cdot B_{\min}}{T_{\text{start}} \cdot F_{\text{drop}} + T_{\text{start}} + 2^{(25 - F_{\min})}} - 5$$

обладает требуемым качеством. Из 129 экспериментов в 128 сравнение показало совпадение результатов обоих используемых подходов оценки качества.

В ходе проведения 66 экспериментов контрольной выборки результаты всех экспериментов, полученные с помощью критерия оценки, совпали с визуальным наблюдением [6].

Полученный критерий оценки качества используется в разработанной системе.

Заключение

Создана система моделирования сетевых помех, которая применяется для эмулирования сетевых параметров и оценки качества образовательных мультимедийных потоков, передаваемых в режиме реального времени в IP-сети. Под оценкой качества в данном случае понимается определение величины битового потока (битрейта), удовлетворяющего конечных пользователей.

Система может использоваться для исследования и тестирования новых мультимедийных сервисов и систем, выявления сетевых проблем и проверки создаваемых критериев оценки качества мультимедийных потоков. Система уже была успешно апробирована при решении проблем, возникших в реальных условиях во время трансляции мультимедийных потоков в сети ПетрГУ.

Как выяснилось, в ходе использования системы наиболее существенное влияние на качество передаваемых мультимедийных потоков оказывают потери сетевых пакетов. При этом, чем больше битрейт мультимедийного потока, тем меньшее количество потерянных пакетов требуется для ухудшения качества видеопотока при заданной ширине сетевого канала.

На основе разработанной системы предполагается создание ПО, реализующего распределенную сеть оценки качества мультимедийных потоков, передаваемых между университетами.

Работа выполнена при финансовой поддержке Программы стратегического развития ПетрГУ в рамках реализации комплекса мероприятий по развитию научно-исследовательской деятельности.

Литература

1. **Забровский А. Л.** Система интерактивного обучения в сети Интернет // Уч. зап. Петрозаводского гос. ун-та. 2009. № 9. С. 63–65.
2. **Linktropy** 5500 WAN Emulator. http://www.apposite-tech.com/products/5500.html (дата обращения: 21.02.2013).
3. **Сетевой** эмулятор WANem. http://wanem.sourceforge.net/ (дата обращения: 21.02.2013).
4. **Видеокодер** WireCast. http://www.telestream.net/wirecast/overview.htm (дата обращения: 21.02.2013).
5. **Strobe** Media Playback. http://www.osmf.org/strobe_mediaplayback.html (дата обращения: 21.02.2013).
6. **Забровский А. Л.** Критерий оценки качества образовательных мультимедийных потоков, транслируемых в реальном режиме времени // Тр. Карельского науч. центра РАН. 2013. № 1. С. 26–32.

УДК 004.434

ФАБРИКИ ПРИКЛАДНОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ, УПРАВЛЯЕМЫЕ МОДЕЛЯМИ ПРЕДМЕТНЫХ ОБЛАСТЕЙ

Н. Д. Андреев,
начальник отдела разработки программного обеспечения
ООО «Джи Джи Эй Софтвэр Сервисес», г. Санкт-Петербург
Ф. А. Новиков,
доктор техн. наук, профессор
Санкт-Петербургский государственный политехнический университет

Обсуждаются методы повышения продуктивности разработки прикладного программного обеспечения на основе определения и использования моделей предметных областей и языков предметной области. Предлагаются принципы применения разработки, управляемой моделью предметной области, для создания фабрик прикладного программного обеспечения. Приводится пример разработанного языка предметной области и указываются преимущества, которые дает его использование.

Ключевые слова *— разработка программного обеспечения, фабрики программного обеспечения, управляемая моделью разработка, предметно-ориентированные языки.*

Введение

В последние годы продолжается совершенствование различных подходов и инструментов для повышения эффективности и предсказуемости разработки прикладного программного обеспечения (ППО). В основе этих подходов лежат методы более эффективного повторного использования накопленных знаний и артефактов, а также повышение уровня абстракции, на котором ведется разработка. Обычно это достигается обобщением решаемых задач, их стандартизацией и последующей автоматизацией. Кроме того, оптимизируется и стандартизируется процесс разработки ППО. Таким образом, создаются фабрики по разработке однородного ППО [1].

Для повышения уровня абстракции обычно предлагается использовать проблемно-ориентированные и предметно-ориентированные языки [2]. Проблемно-ориентированные языки нужны для решения конкретной проблемы или ряда сходных проблем в различных предметных областях. Предметно-ориентированные языки имеют специфику конкретной предметной области [3]. Граница между проблемно- и предметно-ориентированными языками достаточно условна, и в большинстве случаев можно использовать обобщающий термин: язык предметной области [4].

Кроме того, для многих проектов разработки ППО разумно повышать уровень абстракции за счет ведения части или всей разработки на уровне модели создаваемого ППО. В этом случае говорят, что разработка управляется моделью. При этом оказывается, что многие модели проще и удобнее конструировать и поддерживать визуально [5]. Общепринятым средством визуального моделирования в настоящее время является унифицированный язык моделирования UML [6].

Этот же метод разработки лежит в основе инициативы MDA (*Model Driven Architecture* — архитектура, управляемая моделью) [7]. Основная идея этого подхода — создание платформенно независимой модели и ее автоматическое преобразование в целевые платформы.

В данной работе предлагаются подходы к созданию фабрик ППО, управляемых моделями. В качестве примера рассматривается фабрика по созданию интеграционных веб-сервисов.

Подходы к построению фабрик ППО, управляемых моделями предметных областей

Центральным понятием в нашем подходе является модель предметной области (МПО). В целом модель предметной области применительно к разработке ППО — это абстрактное представле-

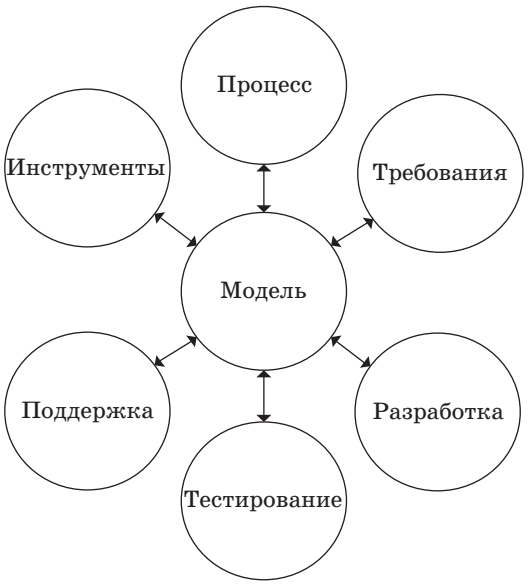
ние знаний, которые определяют множество программно решаемых задач и способов их решения в данной предметной области. В нашем подходе МПО состоит из статического и динамического представлений. Статическое представление является конечным множеством сущностей и множеством отношений между сущностями, что удобно описывается (ориентированным) графом, в котором узлами являются сущности предметной области, а дугами (при необходимости, мультидугами и гипердугами) — отношения между сущностями. Динамическое представление МПО является множеством операций над сущностями и композиций этих операций, т. е. алгоритмов.

Наш многолетний опыт [8] показывает, что моделирование предметной области занимает центральное место в разработке ППО, причем качество МПО существенно влияет на продуктивность разработки. При этом статическое представление модели имеет большее значение, чем динамическое представление.

Моделирование предметной области при разработке ППО является необходимым по существу: невозможно разработать программное обеспечение случайно, не понимая, что и как программное обеспечение должно делать. Но на практике иногда бывает, что моделирование предметной области остается неявной частью процесса разработки, т. е. МПО не разрабатывается в явном отчуждаемом виде, а остается в головах у разработчиков. Это влечет за собой дублирование и одновременно рассогласование моделей в различных артефактах. Например, неявная модель в коде может отличаться от неявно заданной модели в описании требований к ППО, что приводит к несоответствию требованиям и снижению качества ППО.

Основа нашего подхода — явное выделение моделирования предметной области как центральной части разработки ППО. Взаимосвязи МПО с основными составляющими разработки ППО показаны на рис. 1.

1. Требования. Модель предметной области, несомненно, является основой требований к ППО, а конструирование модели — первым шагом в процессе сбора требований и их анализа. Если создается продукт с некоторым шаблонным и стандартным поведением, то, по существу, МПО и является требованиями к ППО. Кроме того, модель является основным связующим звеном между пользователями и командой разработки ППО. Действительно, хорошая модель достаточно абстрактна, не содержит технических деталей и поэтому понятна пользователям. В то же время модель обладает ясной семантикой, а потому является важным источником информации для команды разработки.



■ ***Рис. 1.*** Разработка ППО, управляемая моделью предметной области

2. Разработка. Модель предметной области является основным словарем для разработчиков. Она определяет не только программный код для операций над сущностями и способы хранения сущностей, но может определять, например, структуру и поведение пользовательского интерфейса, интеграцию с другими системами и подсистему верификации данных.

3. Тестирование. Основу тестирования составляют входные и выходные данные системы — разумный перебор входных и верификация выходных данных. МПО определяет свойства данных и, таким образом, во многом определяет набор тестовых примеров.

4. Поддержка. Под поддержкой ППО обычно подразумеваются исправление ошибок после выпуска системы и внесение локальных исправлений внутри системы. Обычно такими исправлениями являются небольшие изменения бизнес-логики или МПО. Например, в системе документооборота может потребоваться дополнительное свойство документа, что повлечет за собой изменение модели.

5. Инструменты. Модель предметной области обычно создают, используя один из инструментов моделирования. Учитывая взаимосвязь модели со всеми шагами разработки ППО, необходимо интегрировать инструмент моделирования с другими используемыми инструментами. Например, в некоторых случаях, исходя из МПО, можно автоматически создавать задачи разработки ППО в системе управления проекта, т. е. интегрировать эту систему с инструментом моделирования. Или, например, автоматически генериро-

вать программный код по модели для последующего редактирования в системе разработки.

6. Процесс. Какой процесс разработки ППО является наиболее эффективным для конкретного проекта, зависит от множества факторов. Тем не менее, по нашему мнению, решающим фактором является стоимость внесения изменений в систему на различных этапах разработки. Применение подхода разработки, управляемой моделью, позволяет минимизировать дублирование и рассогласование знаний, что в свою очередь позволяет минимизировать стоимость внесения изменений в систему.

Важно отметить, что, несмотря на тесную взаимосвязь МПО с различными этапами разработки, в большинстве случаев МПО пока не может полностью определять все артефакты разработки. К сожалению, иногда идеологи разработки, управляемой моделью, настаивают на том, что система всегда должна полностью определяться моделью. Мы считаем, что главной целью должна оставаться эффективность разработки ППО в целом.

Фабрики программного обеспечения позволяют создавать ряд однородных продуктов одинаковым образом и таким образом повышать повторное использование идей, концепций, знаний и артефактов. Процесс создания и использования фабрики ППО, управляемой МПО, представлен на рис. 2.

На первом шаге при создании фабрики программного обеспечения необходимо определить границы обобщения, т. е. тип продуктов, для которого фабрика будет предназначена. По сути это задает круг задач, которые фабрика позволяет решать. При этом можно выбрать два ортогональных направления обобщения:

— проблемный — например, фабрика предназначена для создания сервисов;

— предметный — например, фабрика создается для приложений в области банковского бизнеса.

Далее необходимо определить стандарты и выделить повторяющиеся задачи в рамках фабрики. Все это определяет требования к языку моделирования, проектируемому и реализуемому на следующем шаге. Если мы создаем фабрику проблемного уровня, то создаваемый язык будет тоже проблемно-ориентированным. Если мы создаем фабрику предметного уровня, то создаваемый язык будет тоже предметно-ориентированный.

Если не повышать уровень абстракции за счет создания языка и при этом следовать принятым стандартам и выполнять повторяющиеся задачи вручную, то в рамках фабрики будет расти дублирование знаний и кода. С помощью созданного языка можно определять МПО продуктов с семантикой, которая позволит минимизировать дублирование в рамках фабрики, т. е. создавать часть артефактов разработки автоматически.

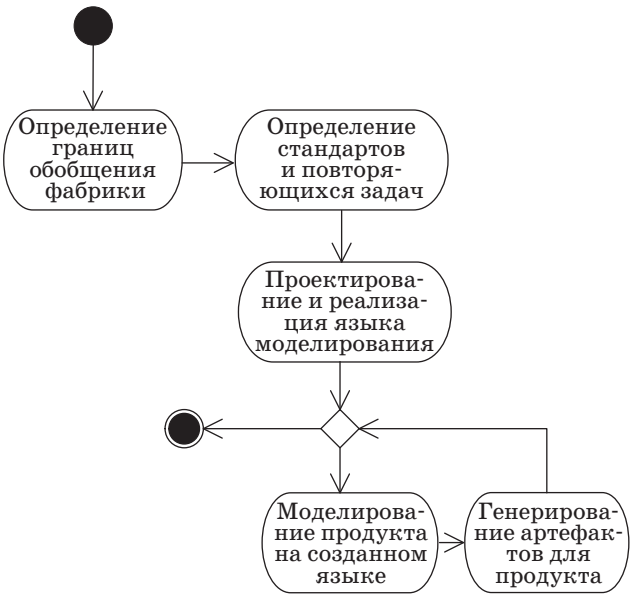
Важно построить эффективный процесс по созданию и развитию фабрик ППО. Мы считаем, что тут необходим инкрементальный процесс [3]. В этом случае создаются одна или несколько фабрик, поддерживающих разработку на проблемных или предметных уровнях. Если при этом использовать подход разработки ППО, управляемой моделью, то следует создать проблемно- или предметно-ориентированные языки моделирования.

Мы предлагаем следующие принципы совместного использования фабрики ППО и разработки, управляемой МПО.

• Явное выделение моделирования предметной области как центральной части разработки ППО. При построении МПО различаются статическое и динамическое представления, причем статическое представление описывается графом, задающим отношения между сущностями. В большинстве случаев граф МПО разумно представлять и изменять визуально.

• Создание проблемно-ориентированного языка моделирования предметной области. Синтаксис и метамодель языка моделирования основываются на языке UML [9]. Созданный язык определяет стандарты и уровень обобщения в рамках фабрики. Семантика созданного языка позволяет избежать дублирования концепций и кода через создание генераторов.

• Модель продукта на созданном проблемно-ориентированном языке предопределяет характеристические свойства продукта, но не довлеет над деталями реализации. Часть артефактов может быть создана традиционными способами.



■ ***Рис. 2.*** Создание и использование фабрики ППО, управляемой МПО

Возможно продолжение разработки на традиционном языке программирования в любой момент.

• Создание языка и последующее его использование осуществляются в специализированном инструменте, который позволяет производить метамоделирование, задавать свойства графического синтаксиса языка и правила, накладываемые на метамодель. Правила задаются на языке OCL [10]. В процессе моделирования на созданном языке инструмент верифицирует модели на соответствие заданным правилам.

Проблемно-ориентированный язык SOALang для описания сервисов

Особенностью программных систем уровня предприятия является долгое время жизни и интеграция систем между собой. Любое изменение программных интерфейсов одной системы необходимо проводить вместе с изменениями зависимых систем. При большом количестве систем и связей между ними подобные изменения являются рискованными и дорогостоящими. Сервис-ориентированная архитектура (*Service-Oriented Architecture* — SOA) [11] призвана решить эти проблемы посредством интеграции систем через слабо связанные между собой сервисы. Такие сервисы не имеют прямых вызовов друг друга, и системы интегрируются не напрямую, а через сервисную шину предприятия [12]. При этом программные интерфейсы и МПО, выставляемые потребителю интеграционных сервисов, должны быть каноническими, т. е. не зависеть от используемых систем и деталей реализации сервисов.

Таким образом, при изменении или замене используемых систем интеграционные сервисы скрывают эти изменения от других систем. Подобные сервисы должны создаваться в соответствии с разработанными и принятыми стандартами, чтобы разработчикам зависимых систем было привычно и удобно пользоваться их интерфейсами и чтобы снизить стоимость поддержки как самих сервисов, так и использующих их систем. Особенно важно следовать этим подходам в крупных компаниях, где количество интегрируемых между собой систем может исчисляться тысячами. При этом разработку интеграционных сервисов разумно обобщить и выполнить в рамках фабрики. Это даст возможность автоматизировать повторяющиеся действия и построить эффективный процесс разработки. Все принятые стандарты необходимо автоматически верифицировать, если соответствующие действия не могут быть автоматизированы и выполняются вручную. Одним из самых важных этапов при разработке интеграционных сервисов является разработка канонической МПО [13].

Для решения вышеописанных задач в компании «Джи Джи Эй» [14] был разработан проблемно-ориентированный язык SOALang, который позволяет задавать модель сущностей, их свойства и связи. Таким образом, язык описывает явно только статическую составляющую интеграционных сервисов.

Язык SOALang основан на подмножестве графического синтаксиса, метамодели и семантики диаграммы классов языка UML. Упрощенная метамодель языка SOALang представлена на рис. 3. При этом введены дополнительные правила и более строгие ограничения, в частности:

1) имя сущности должно быть уникальным существительным в единственном числе;

2) именование сущности должно быть в соответствии со стандартом UpperCamelCase [15];

3) поле сущности должно иметь уникальное имя в рамках сущности и разрешимый тип;

4) именование полей сущности должно быть в соответствии со стандартом lowerCamelCase [15];

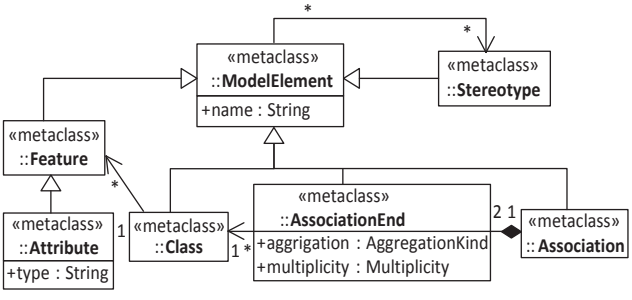
5) имя роли ассоциации должно быть задано, если роль не является агрегацией, композицией или направленной; должно быть во множественном или в единственном числе — в зависимости от свойства множественности роли ассоциации;

6) для каждой сущности должен быть задан один уникальный неизменяемый ключ — задается как стереотип «*unique immutable*» для одного из полей;

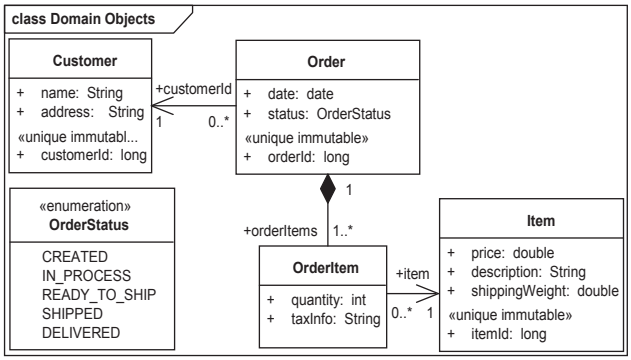
7) у сущности типа Перечисление (стереотип «*enumeration*») все значения должны состоять из прописных букв.

Многие из дополнительных правил задают стандарты именования сущностей и полей. Это необходимо, чтобы выводимые из моделей артефакты были согласованы и похожи между собой в рамках фабрики. Принятые единые стандарты и согласованность интерфейсов в рамках фабрики упрощают использование сервисов в целом.

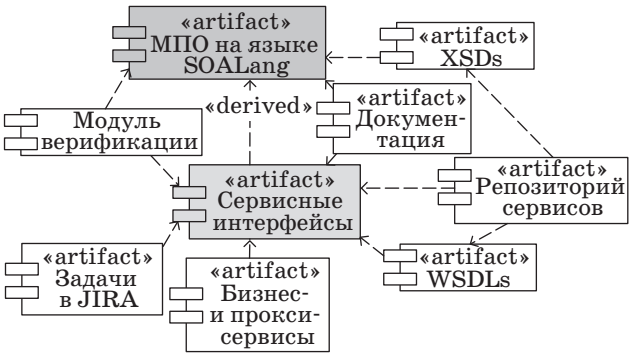
Пример модели, созданной с помощью языка SOALang, показан на рис. 4. В этой модели описывается предметная область ведения базы заказов. В данной МПО выделяются сущности: заказ (*Order*), заказчик (*Customer*), позиция заказа



■ ***Рис. 3***. Упрощенная метамодель языка SOALang



■ ***Рис. 4.*** Пример МПО на языке SOALang



■ ***Рис. 5.*** Язык SOALang и производные артефакты

(*OrderItem*) и товар (*Item*). Сущности имеют атрибуты и связаны отношениями, семантика которых определяется семантикой использованных отношений UML.

В качестве инструмента моделирования используется Enterprise Architect (EA) [16]. Для него можно разрабатывать расширения, в коде которых можно осуществлять доступ к открытым в инструменте моделям — создавать новые или изменять существующие модели. Были разработаны нижеследующие расширения для поддержки разработанного языка SOALang и верификации модели на соответствие правилам, определенным в языке. На рис. 5 показана диаграмма связи модели, артефактов и модуля верификации. Первичным артефактом является МПО на языке SOALang, разработанная бизнес-аналитиком. В следующем разделе описываются выводимые из МПО артефакты и модуль верификации.

Преимущества использования языка SOALang

Под сервисными интерфейсами мы понимаем набор операций над объектами предметной области, предоставляемых сервисом. При этом операции определенным образом группируются в интерфейсы.

В рамках фабрики сервисов определены стандарты и правила, которым необходимо следовать при разработке сервисных интерфейсов. В частности, для каждой сущности создается отдельный сервисный интерфейс. Например, для сущности с именем Entity создается интерфейс с именем EntityService. Исключение составляют:

— перечисления;

— сущности, которые связаны с другими сущностями ассоциацией композиции, т. е. являются частью другой сущности, например OrderItem (см. рис. 4).

При этом в каждом интерфейсе присутствует стандартный набор методов со следующими правилами именования для сущности с именем Entity:

— получение сущности по ключу (атрибут со стереотипом «*unique immutable*») — findEntityByEntityId;

— получение сущностей по списку ключей — findEntitiesByEntityIds;

— получение всех сущностей — findAllEntities.

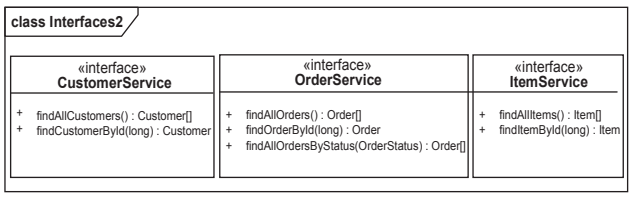
Например, для сущности Order:

— findOrderByOrderId(long orderId) : Order
— findOrdersByOrderIds(long[] orderIds) : Order[]
— findAllOrders() : Order[]

Мы определили эти правила как часть семантики языка SOALang, и это позволило реализовать генератор сервисных интерфейсов по МПО. Таким образом, сервисные интерфейсы и их операции по умолчанию генерируются автоматически как еще одна модель. Перед запуском процесса можно выбрать опцию генерации методов, которые изменяют данные: добавление, изменение и удаление сущности. По умолчанию генерируются только методы чтения, так как обычно интеграционные сервисы только возвращают информацию из различных источников данных, но не записывают ее.

Получившиеся после генерации интерфейсы можно отредактировать вручную. Если будут сделаны изменения модели и опять запущена генерация сервисов, то будут добавлены новые или изменены существующие методы. Добавленные пользователем методы изменены не будут.

Пример сгенерированной модели сервисных интерфейсов из МПО представлен на рис. 6. Стоит отметить, что для сущности OrderItem интер-



■ ***Рис. 6.*** Пример модели сервисных интерфейсов

фейс сгенерирован не был в соответствии с вышеописанными правилами — действительно, сущности Order и OrderItem связаны отношением композиции.

Верификация модели — это проверка ее соответствия требованиям и правилам, определенным для языка. Верификация позволяет определить проблемы моделирования на ранних стадиях. Был реализован модуль верификации МПО и сервисных интерфейсов на соответствие правилам, определенным в языке SOALang. Верификация может быть запущена вручную при редактировании модели и автоматически запускается при генерации артефактов из модели. Мы разделили потенциальные проблемы на две группы:

— ошибка (error) — при таких проблемах дальнейшая генерация невозможна. Например, если имя роли ассоциации не задано, то это ошибка;

— предупреждение (warning) — при таких проблемах будет выдано предупреждение, но работа может быть продолжена. Например, если для какой-то сущности, которая связана отношением композиции с другой сущностью (является частью другой сущности), не будет задан ключ, то будет выдано предупреждение.

В процессе верификации сервисных интерфейсов проверяются правила именования и существование типов, на которые есть ссылки из сигнатур операций.

В используемом инструменте Enterprise Architect верификация правил стандарта UML реализована не полностью. Разработанный нами модуль верификации покрывает пропущенные, но являющиеся частью языка SOALang правила UML, и дополнительные правила языка SOALang. Для верификации модели правила группируют по тем элементам метамодели, к которым правила применимы. Кроме того, для эффективного процесса верификации формируются логические цепочки зависимостей правил [17]. Разработанный модуль верификации делает обход модели и применяет соответствующие правила к каждому элементу.

Файлы стандартов WSDL [18] и XSD [19] описывают интерфейс сервисов для потенциальных потребителей. Для обеспечения совместимости и согласованности были разработаны обязательные стандарты и правила создания WSDL и XSD. Встроенные возможности инструмента Enterprise Architect по генерации этих артефактов не отвечают нашим стандартам. Поэтому мы разработали собственный генератор WSDL и XSD. Например, тип агрегирования в отношении между сущностями влияет на XSD. Если отношение является композицией, то для зависимой сущности будет сгенерирован вложенный элемент. Если отношение является агрегацией, то будет сгенерирована ссылка на зависимую сущность. Ниже представлен фрагмент XSD, сгенерированного по модели:

```
<xs:complexType name="Order">
 <xs:sequence>
  <xs:element name="date" type="xs:date"
              minOccurs="0" maxOccurs="1"/>
  <xs:element name="status" type="tns:OrderStatus"
              minOccurs="0" maxOccurs="1"/>
  <xs:element name="orderId" type="xs:long"
              minOccurs="0" maxOccurs="1"/>
  <xs:element name="orderItems" type="tns:OrderItem"
              minOccurs="1" maxOccurs="unbounded"/>
  <xs:element name="customerId" type="xs:long"
              minOccurs="1" maxOccurs="1"/>
 </xs:sequence>
</xs:complexType>
```

Файлы XSD создаются по МПО, а по сервисным интерфейсам генерируются файлы WSDL. При этом в WSDL добавляются требуемые стандартом параметры безопасности и пространства имен. Для создания и отслеживания статуса задач и ошибок используется система JIRA [20]. Для разрабатываемых сервисов готовятся и проводятся автоматические функциональные и нагрузочные тесты. Таким образом, для каждого метода сервиса существуют следующие стандартные задачи:

1) разработка метода;

2) подготовка функциональных тестов с помощью SOAP UI [21];

3) подготовка тестов производительности/нагрузки с помощью JMeter [22].

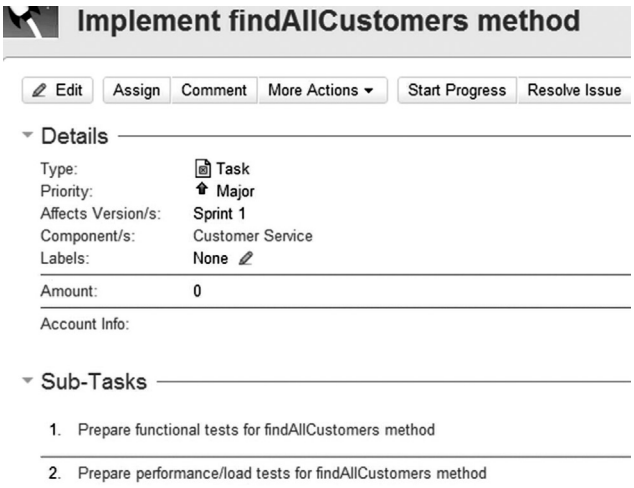
Система JIRA поддерживает удобный программный интерфейс, который, в частности, позволяет добавлять новые задачи. Была реализована подсистема автоматического создания вышеописанных задач в JIRA по сервисной модели. В качестве компоненты указывается соответствующий сервис, а задачи типа 2, 3 (подготовка автоматических тестов) создаются как подзадачи соответствующей задачи на разработку метода сервиса. На рис. 7 показан пример автоматически полученной задачи в JIRA.

Обычно при разработке интеграционных сервисов реализуют следующие дополнительные сервисные прослойки на сервисной шине предприятия:

— бизнес-сервисы — для балансировки нагрузки, кэширования результатов, обеспечения гарантированной доставки;

— прокси-сервисы — для маршрутизации запросов и ответов, трансформации данных, управления транзакциями.

При полноценном использовании сервис-ориентированной архитектуры создание бизнес- и прокси-сервисов является обязательным. При этом на этих уровнях подключаются некоторые стандартные механизмы, например безопасность. Часто получается, что эти прослойки не выпол-



■ *Рис. 7.* Пример автоматически созданной задачи в JIRA

няют никакой дополнительной функции и реализуются стандартным образом. Был реализован генератор XML-файлов, описывающих конфигурацию бизнес- и прокси-сервисов из МПО и сервисных интерфейсов. Ниже приводится фрагмент конфигурации XML для бизнес-сервиса:

```
<ser:binding type="SOAP" isSoap12="false"
  xsi:type="con:SoapBindingType"
  xmlns:con="http://www.bea.com/wli/sb/services/bindings/config"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <con:wsdl ref="OrderingServices/Resources/
    CustomerInteraction_1_0/CustomerInteraction"/>
  <con:port>
    <con:name>CustomerResponderPort</con:name>
    <con:namespace>
      urn:namespace:services:Customer:1.0
    </con:namespace>
  </con:port>
</ser:binding>
```

Таким образом, мы уменьшаем дублирование и гарантируем использование принятых стандартов.

Частью процесса разработки интеграционных сервисов является подготовка документации. Был реализован генератор документов по МПО и сервисных интерфейсов. Получаемая документация не является полной, и ее необходимо дополнять вручную. Тем не менее, генератор существенно упрощает создание документации и помогает следовать принятым шаблонам.

Для эффективного использования интеграционных сервисов в крупных компаниях необходим репозиторий сервисов, где собирается информация о свойствах, версиях и возможностях сервисов, взаимосвязях между собой и зависимостях от внешних систем. Была реализована интеграция с репозиторием SOA Lifecycle Manager [23] для автоматического добавления туда информации о сервисах по модели, включая описания WSDL.

Заключение

В работе рассмотрен пример фабрики по созданию интеграционных веб-сервисов, управляемой МПО. На примере показано, что если в рамках фабрики определить стандарты и правила для получаемых артефактов, то многие процессы можно оптимизировать. При этом центральным элементом становится МПО. На основе обобщения этого и других примеров предложены принципы построения фабрик ППО, управляемых моделями.

Дальнейшее развитие работы мы видим в определении подходов к декларативному описанию операционной семантики. Например, указывать дополнительные свойства на МПО, которые позволят генерировать часть программного кода. При этом важно сохранить платформенно-независимую модель. Для рассмотренного примера интеграционных сервисов это даст возможность генерировать часть кода сервисов, например на языке BPEL или Java.

Литература

1. **Greenfield J., Short K., Cook S., Kent S.** Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools. — Wiley, 2004. — 500 p.
2. **Dmitriev S.** Language Oriented Programming: The Next Programming Paradigm. http://www.onboard.jetbrains.com/articles/04/10/lop/mps.pdf (дата обращения: 08.03.2012).
3. **Андреев Н. Д., Новиков Ф. А.** Инкрементальный предметно-ориентированный процесс разработки прикладного программного обеспечения // Информационно-управляющие системы. 2012. № 1. С. 60–66.
4. **Новиков Ф. А., Тихонова У. Н.** Автоматный метод определения проблемно-ориентированных языков // Информационно-управляющие системы. 2009. № 6. С. 34–40; 2010. № 2. С. 31–37; № 3. С. 29–37.
5. **Новиков Ф. А.** Визуальное конструирование программ // Информационно-управляющие системы. 2005. № 6. С. 9–22.
6. **Новиков Ф. А., Иванов Д. Ю.** Моделирование на UML. Теория, практика, видеокурс. — СПб.: Наука и Техника, 2010. — 640 с.

7. **Booch G.** et al. An MDA Manifesto // The Mda J.: Model Driven Architecture Straight From The Masters. Meghan Kiffer Press, Dec. 2004. Chap. 11. P. 2–9.
8. **Новиков Ф. А.** Методы алгоритмизации предметных областей: дис. ... д-ра техн. наук. СПб.: СПбГУ ИТМО, 2011. http://www.dissercat.com/content/metody-algoritmizatsii-predmetnykh-oblastei (дата обращения: 08.03.2012).
9. **Андреев Н. Д.** Предметно-ориентированный язык моделирования, основанный на UML // Формирование технической политики инновационных наукоемких технологий: материалы конф. и школы-семинара / СПбГПУ. СПб., 2004. С. 75–82.
10. **Язык** OCL. http://en.wikipedia.org/wiki/Object_Constraint_Language (дата обращения: 08.03.2012).
11. **Service-Oriented** Architecture. http://en.wikipedia.org/wiki/Service-oriented_architecture (дата обращения: 08.03.2012).
12. **Enterprise** Service Bus. http://en.wikipedia.org/wiki/Enterprise_service_bus (дата обращения: 08.03.2012).
13. **Hohpe G., Woolf B.** Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions. — Addison-Wesley, 2003. — 736 p.
14. **http:**//www.ggasoftware.com/ (дата обращения: 08.03.2012).
15. **Camel** case. http://ru.wikipedia.org/wiki/CamelCase (дата обращения: 08.03.2012).
16. **Enterprise** Architect. http://www.sparxsystems.com/ (дата обращения: 08.03.2012).
17. **Андреев Н. Д.** Автоматическая верификация модели UML // IV Междунар. молодежная школа-семинар «БИКАМП-2003»: тр. конф. / ГУАП. СПб., 2003. С. 145–149.
18. **WDL.** http://www.w3.org/TR/wsdl (дата обращения: 08.03.2012).
19. **XML** Schema. http://www.w3.org/XML/Schema.html (дата обращения: 08.03.2012).
20. **Инструмент** JIRA. http://www.atlassian.com/software/jira/overview (дата обращения: 08.03.2012).
21. **Инструмент** SOAP UI. http://www.soapui.org (дата обращения: 08.03.2012).
22. **Инструмент** JMeter. http://jmeter.apache.org/ (дата обращения: 08.03.2012).
23. **Продукт** Lifecycle Manager. http://www.soa.com/products/lifecycle_manager (дата обращения: 08.03.2012).

ПАМЯТКА ДЛЯ АВТОРОВ

Поступающие в редакцию статьи проходят обязательное рецензирование.

При наличии положительной рецензии статья рассматривается редакционной коллегией. Принятая в печать статья направляется автору для согласования редакторских правок. После согласования автор представляет в редакцию окончательный вариант текста статьи.

Процедуры согласования текста статьи могут осуществляться как непосредственно в редакции, так и по e-mail (80x@mail.ru).

При отклонении статьи редакция представляет автору мотивированное заключение и рецензию, при необходимости доработать статью — рецензию. Рукописи не возвращаются.

Редакция журнала напоминает, что ответственность за достоверность и точность рекламных материалов несут рекламодатели.

УДК 681.3.06 (075.8)

ВИЗУАЛЬНЫЙ АНАЛИЗ ЗАЩИЩЕННОСТИ КОМПЬЮТЕРНЫХ СЕТЕЙ

И. В. Котенко,
доктор техн. наук, профессор, заведующий лабораторией проблем компьютерной безопасности
Е. С. Новикова,
канд. техн. наук, старший научный сотрудник
Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН)

Исследуются методики визуального анализа защищенности компьютерных сетей. Описывается компонент визуализации системы оценки защищенности компьютерной сети, отличающийся от других систем тем, что позволяет графически представлять как отчеты сканеров уязвимостей, так и результаты моделирования атак, благодаря чему пользователь системы может соотнести потенциальные причины нарушения безопасности с возможными последствиями их эксплуатации злоумышленником.

Ключевые слова *— визуализация событий безопасности, оценка защищенности, политики безопасности, графы атак, карты деревьев.*

Введение

Методики визуального анализа данных позволяют эффективно исследовать данные большого объема и извлекать новые знания из массива неоднородных, зашумленных данных. Основная идея визуальной аналитики заключается в объединении особенностей зрительного восприятия человеком информации и мощностей электронной обработки данных, в результате чего возможно создание высокоинтерактивного программного обеспечения, позволяющего пользователю погрузиться в данные, лучше понимать результаты алгоритмов их обработки и вести процесс исследования в наиболее перспективном направлении [1].

Методики визуального анализа широко используются для анализа безопасности информационной системы. В настоящее время большая часть существующих решений предназначена для эффективного контроля периметра сети [2, 3]. Имеются различные инструменты для анализа состояния всей сети в целом, мониторинга портов и определения различных паттернов сканирования портов, выявления аномалий в «сетевом поведении» пользователя, в то время как вопросы визуализации данных об уровне защищенности компьютерной сети, поддержки принятия решений проработаны в меньшей степени [2, 3].

В настоящей работе представлены модели и методики визуального анализа, реализованные в системе оценки защищенности компьютерных сетей [4–6], которая позволяет графически определить наиболее уязвимые места информационной системы, сформировать шаблоны атак в зависимости от начальных условий атак и на основе полученных данных соответствующим образом скорректировать план мероприятий по обеспечению безопасности системы. Главным отличием представляемой системы является возможность визуально анализировать потенциальные причины в контексте возможных последствий их эксплуатации.

Методики визуализации для анализа защищенности компьютерных сетей

Механизмы визуализации, предназначенные для анализа защищенности сети и поддержки принятия решений администратором сети, представлены в научных работах не столь широко. Кроме того, иногда сложно провести четкую границу между инструментами визуализации исходя из области их применения. Так, например, систему SpiralView, предназначенную для поддержки принятия решений системным администратором, успешно можно применять для мониторинга сетевого трафика [7], поскольку в ней используется визуализация событий безопасности, регистрируемых различными датчиками безопасности, в том числе и системными утилитами, фиксирующими действия пользователей и при-

ложений, в режиме реального времени. Для графического представления информации используется подход, предложенный в работе [8]: события располагаются на окружностях, радиус которых является шкалой времени (рис. 1).

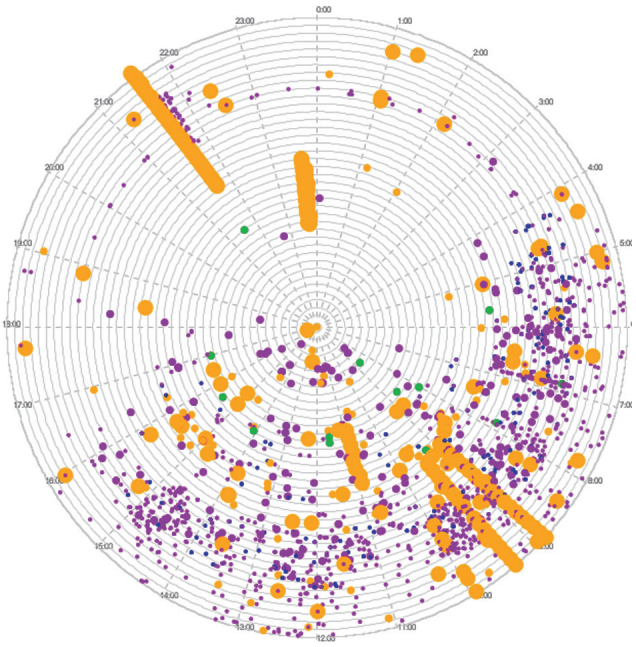
Тип событий маркируется цветом, и пользователь имеет возможность отфильтровать или выделить цветом данные в соответствии с заданными им условиями. Однако такая модель визуализации информации не представляется удобной для оценки корректности используемых в системе политик безопасности, поскольку помогает оператору выявить небезопасные элементы системы, но для понимания причин их появления требуется выполнение дополнительного анализа.

Для визуального анализа политик безопасности межсетевых экранов Тран и др. [9] разработали инструмент PolicyVis, который отображает правила межсетевого экрана в виде прямоугольников. Положение и геометрия прямоугольника определяются тремя полями правила, выбираемыми пользователем. Цветом кодируется статус трафика (зеленый — разрешенный трафик, красный — блокируемый трафик). Благодаря такому представлению пользователь может легко выявить различные аномалии в политике безопасности (избыточность, затенение, обобщение, корреляцию), о которых свидетельствуют пересекающиеся прямоугольники.

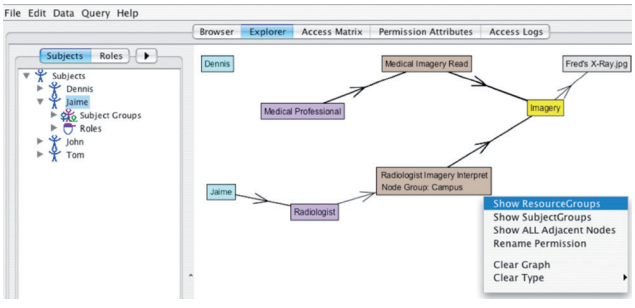
Мансманн и др. [10] для визуализации политик безопасности адаптировали модель визуализации «солнечные лучи» (Sunburst), которая позволяет компактно графически представлять иерархическую структуру. Корневой элемент структуры помещается в центр в виде круга, а элементы каждого уровня иерархии рекурсивно отображаются на соответствующие сегменты кольца. Для использования данной модели визуализации авторы разработали правила преобразования политики безопасности в иерархическую структуру. Согласно им, первый уровень после корневого узла состоит из названий различных списков контроля доступа, второй уровень содержит описания прав доступа («разрешить» или «запретить»), на третьем уровне располагаются названия протоколов («tcp», «ip», «udp» и т. д.), за которыми следуют IP-адреса получателей и отправителей.

Интересные результаты можно получить при представлении списков доступа пользователей к ресурсам в виде связных графов, вершины которых соответствуют пользователям/группам пользователей и информационным ресурсам, а ребра обозначают возможность получения доступа к объекту [11, 12]. Цветом обычно обозначаются роли пользователя/группы пользователей. Например, инструмент RubaViz [12] использует два графических представления правил доступа к ресурсам: матричное и в виде графа. Пример графа, соответствующего правилам доступа, представлен на рис. 2 [12]. Р. Марти [11] показал, что такое графическое представление в сочетании с алгоритмами кластеризации и компоновки графа, учитывающими его семантику, позволяет сформировать как стандартные модели поведения пользователей, так и отклонения от них.

Хайнтцман и др. [13] предложили представлять права доступа к файловым ресурсам в стандартной иерархической файловой системе в виде карты деревьев, вложенные прямоугольники которой соответствуют файлам и папкам. С помощью цвета кодируются их разрешения, т. е. узел карты деревьев, соответствующий заданной папке или файлу, изображается зеленым, красным или серым, если разрешения к нему слабее, сильнее или равны базовому значению соответственно, которое может принимать следующие значения: «нет доступа», «чтение», «чтение и запись» и «полный доступ». Кроме того, узлы карты дерева



■ *Рис. 1.* Модель визуализации политик безопасности



■ *Рис. 2.* Представление правил доступа к ресурсам в виде графа

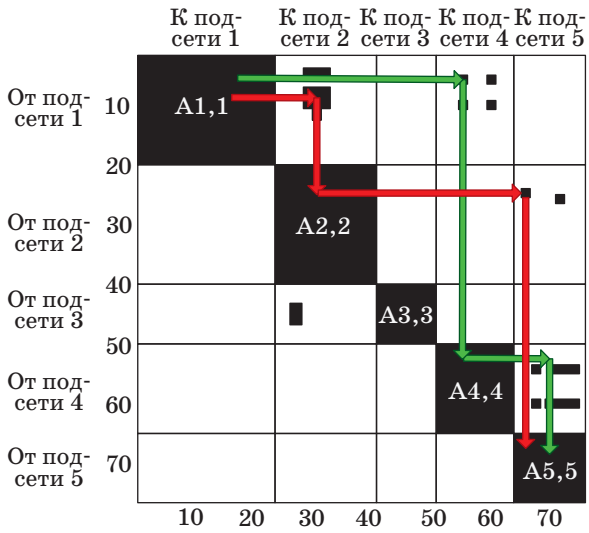
выделяются оранжевой рамкой, если имеет место нарушение разрешений родительского узла.

Карты деревьев широко применяются для анализа выявленных уязвимостей в компьютерных сетях [11, 14]. Например, веб-приложение Nv [14] представляет отчеты сканера уязвимостей Nessus [15] в виде карт деревьев и гистограмм. Помимо графической интерпретации результатов одного сканирования инструмент позволяет оценить прогресс в устранении обнаруженных уязвимостей, показывая, какие уязвимости были устранены, какие остались неразрешенными и какие новые уязвимости появились в системе. В инструменте используется семантическая цветовая схема, в рамках которой устраненные уязвимости обозначаются зеленым цветом, новым уязвимостям соответствует красный цвет, а оранжевым обозначаются уязвимости, находящиеся в работе.

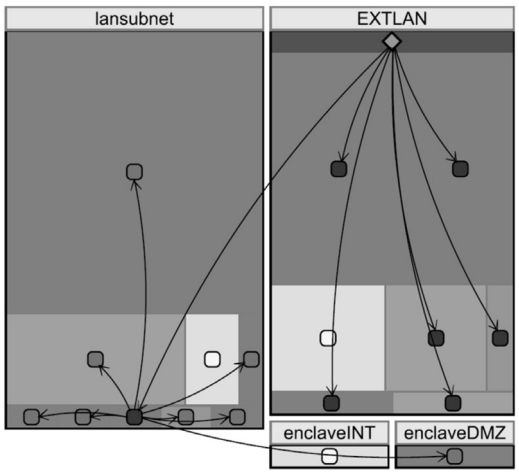
Графы атак являются одним из способов анализа защищенности сети. Графы атак — важный инструмент для оценивания уровня защищенности сети и выявления потенциальных путей проникновения в систему злоумышленником [4–6, 16]. Естественным представлением результатов моделирования атак являются сами графы. Вершинами графа являются различные хосты сети и уязвимости, эксплуатируемые злоумышленником по мере продвижения от одной скомпрометированной машины к другой, а дуги отражают порядок выполнения действия атакующего. Однако, как показано в работе [17], сложность графа атак квадратично зависит от числа хостов в анализируемой сети, поэтому в большинстве случаев традиционное представление графов нечитаемо из-за большого количества узлов и связей между ними.

Для анализа возможных шагов злоумышленника предлагается [17] использовать матрицы смежности, которые являются альтернативным способом представления графов. Ненулевой элемент матрицы a_{ij} обозначает дугу между i-й и j-й вершинами графа атак (рис. 3). Ряды и столбцы матрицы могут быть упорядочены любым образом, при этом структура графа атак остается неизменной. С помощью такого графического представления уменьшается сложность анализируемых данных, кроме того, можно пошагово отследить развитие атаки, выделить определенные шаблоны атак и классифицировать их в зависимости от исходных условий.

В работе [18] предложен способ представления графов атак, который позволяет спроецировать результаты моделирования атаки на физическую топологию сети. Каждая подсеть представляется в виде карты деревьев, вложенные прямоугольники которой символизируют узлы, с помощью цвета кодируются различные атрибуты узлов, а размер пропорционален числу скомпрометированных узлов в подсети (рис. 4) [18]. Этот подход



■ ***Рис. 3.*** Представление графа атак в виде матрицы смежности



■ ***Рис. 4.*** Представление графов атак в виде карты деревьев

реализован в системе Navigator [19]. Пользователь имеет возможность располагать карты деревьев в произвольном порядке, чтобы получить интуитивно понятный вид топологии исследуемой сети. Кроме того, инструмент позволяет проводить эксперименты вида «что-если», благодаря этому администратор сети может оценить необходимость установки различных патчей, изменения правил межсетевых экранов и т. д.

Модели и методики визуального анализа, применяемые в системе оценки защищенности компьютерной сети

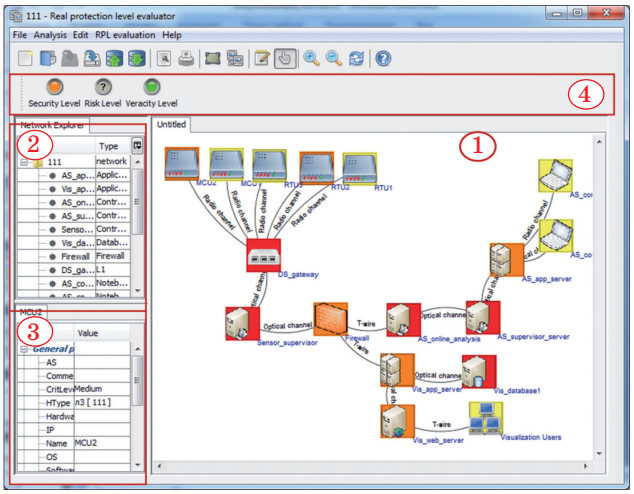
Система оценки защищенности компьютерной сети позволяет оценить уровень ее защищенности, основываясь на результатах аналитического

и динамического моделирования атак и расчета метрик безопасности [4–6]. Графический интерфейс пользователя предоставляет оператору системы возможности по конфигурированию исходных данных и представлению результатов моделирования атак в графическом виде.

Главное окно системы разделено на четыре функциональных вида (рис. 5).

Главный вид *1* представляет топологию исследуемой сети в виде графа, в то время как вид *2* отражает иерархическую структуру, показывая домены или группы хостов. Пользователь может добавлять и удалять узлы компьютерной сети. Пиктограммы оборудования являются настраиваемыми, поэтому пользователь может задать иконку для обозначения типа сетевого объекта. Фон пиктограммы используется для отображения значений метрик безопасности, вычисленных в результате работы системы оценки защищенности компьютерной сети. Пользователь может выбрать метрику из предопределенного списка {Уровень критичности, Уровень риска, Ущерб, Число уязвимых приложений}. Краткая информация по каждому хосту также доступна в виде всплывающей подсказки, которая появляется при наведении указателя мышки пользователем на объект сети. Свойства узлов сети задаются при помощи редактора свойств *3*, причем пользователь может сконфигурировать каждый узел сети и саму сеть в целом. Он может задать как значения заранее определенных свойств, таких как IP-адрес, тип хоста (веб-сервер, файловый сервер, роутер и т. д.), установленное программное и аппаратное обеспечение, так и определить собственные свойства хостов.

Панель управления *4* используется для отображения значений метрик защищенности: она отображает уровень защищенности анализируемой сети, уровень рисков, уровень достоверности информации. Такое расположение информации об исследуемой сети позволяет оценить ее состояние в целом, и пользователь имеет возможность проанализировать результаты оценки защищенности сети в контексте исходной информации, которая доступна в разнообразных видах, расположенных на одной панели управления.



■ *Рис. 5.* Главное окно системы анализа защищенности компьютерной сети

Для представления компьютерных сетей большого размера используется обычное геометрическое и семантическое масштабирование. Применение механизмов семантического масштабирования позволяет проводить агрегирование узлов графа, исходя из значений свойств узла (принадлежность к рабочей группе, домену и т. д.). Агрегирование узлов происходит в интерактивном режиме: пользователь может свернуть/развернуть часть сети, выбирая соответствующий пункт контекстного меню выбранного узла сети.

Отличительной особенностью разрабатываемой системы является возможность визуально анализировать отчеты сканеров уязвимости и графов атак одновременно; таким образом, пользователь системы может оценить потенциальные причины нарушения безопасности в компьютерной сети и возможные последствия их эксплуатации злоумышленником.

Выявленные уязвимости оцениваются при помощи метрик, определенных системой CVSS [20]. Статистическая информация по выявленным уязвимостям представляется с помощью простых графических моделей, применяющих секторные и пузырьковые диаграммы.

Например, секторные диаграммы используются для отображения распределения уязвимостей с учетом их критичности (Severity), сложности реализации (Access Complexity), уровня ущерба (Mortality) для одного хоста и для всей сети в целом. При этом пользователь может выбрать сектор диаграммы и, нажав на него мышью, получить перечень уязвимостей, попавших в заданную категорию. Информация о наиболее часто встречаемых уязвимостях в системе и наиболее уязвимых хостах также представляется в виде секторной диаграммы.

Пузырьковые диаграммы используются для анализа сложности реализации и критичности уязвимостей, выявленных на одном хосте, т. е. пользователь имеет возможности оценить число наиболее критичных уязвимостей в контексте сложности их эксплуатации.

Поскольку указанные метрики могут принимать ограниченное число значений: {Высокий, Средний, Низкий} для показателя критичности и {Высокий, Низкий} для сложности их эксплуатации, — то число комбинаций этих значений равно шести, благодаря чему генерируемое изображение не перегружено и легко читается поль-

зователем. Размер пузырька определяется числом уязвимостей, принадлежащих заданной категории, а цвет определяется комбинацией этих двух метрик: чем выше уровень критичности и ниже уровень сложности эксплуатации уязвимости, тем насыщенней и ярче красный цвет.

Данные графические модели представления просты и понятны пользователю, кроме того, могут быть использованы в отчетной документации любого уровня. Однако они не подходят для представления данных большого объема, так как оптимальное число различных категорий данных, отображаемых с их помощью, равно 10–15 [11]. Для формирования общего представления о выявленных уязвимостях в системе используются карты деревьев, каждый вложенный прямоугольник которых обозначает узел сети. Используя такие атрибуты, как размер прямоугольника и его цвет, можно закодировать атрибуты анализируемого объекта. Например, на рис. 6 размер прямоугольников определяется уровнем критичности узла, назначаемым пользователем. Таким образом, наиболее важные с точки зрения пользователя узлы более заметны. Цвет используется для обозначения критичности выявленных на хосте уязвимостей. Так, красный цвет соответствует высокому уровню критичности, а желтый — низкому уровню.

Такой подход позволяет определить или изменить план мероприятий для повышения уровня защищенности компьютерных сетей, например график обновлений и замены программного обеспечения. Для того чтобы позволить пользователю работать с крупномасштабными сетями, предусмотрен гибкий механизм масштабирования, позволяющий отображать выбранную часть (домен, рабочую группу) карты деревьев, определяемую иерархией сети.

Следует отметить, что для представления результатов анализа уязвимостей мы используем семантическую цветовую схему при представлении значений метрик — от желтого к красному. Зеленые цвета в отчетах об уязвимостях не используются, так как они обычно применяются для обозначения нормальных (безопасных) значений показателей, а любая уязвимость потенциально несет угрозу для безопасности системы.

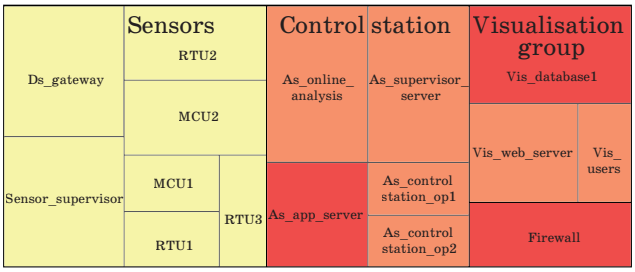
Для представления результатов моделирования атак используются графы. Каждый узел графа соответствует определенному атакующему действию, а их порядок отражает последовательность действий, выполняемых злоумышленником: узлы, расположенные на одном уровне, обозначают действия, которые могут быть выполнены одновременно или независимо друг от друга, а узлы, расположенные на разных уровнях, обозначают действия, которые выполняются в определенной последовательности.

В системе используются условные обозначения (таблица). Для обозначения типа действия применяются одновременно цвет и форма пиктограммы, благодаря этому с помощью цвета могут быть закодированы характеристики, вычисленные для каждого действия.

Графы атак позволяют изучить развитие атаки в анализируемой компьютерной сети, отслеживая действия нарушителя. Однако исследования показали, что графы атак могут обладать высокой степенью связности и быть чрезвычайно сложными, что значительно затрудняет их применение на практике [17]. В целях упрощения и повышения эффективности процесса их анализа были разработаны следующие способы взаимодействия с графическим представлением графов атак.

Геометрическое масштабирование. Позволяет пользователю сфокусироваться на определенных частях графа и уменьшить уровень связности графа. Расстояние между узлами графа может быть динамически изменено с помощью колесика мыши.

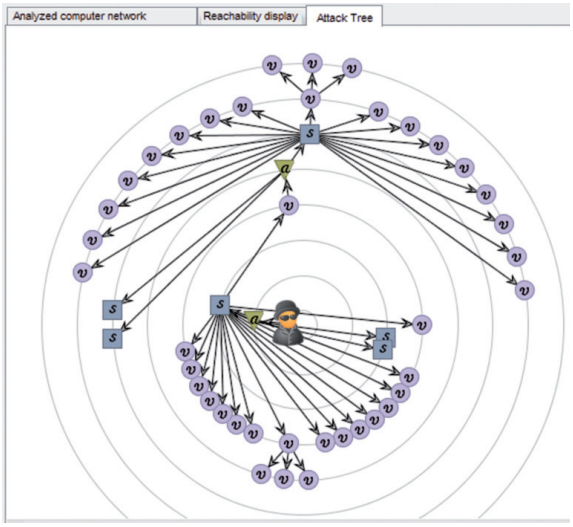
Настройка компоновки графа. В настоящее время в системе поддерживаются два алгоритма компоновки графа — древовидный и радиальный. Радиальное расположение графа более компактно и позволяет пользователю увидеть граф атак целиком (рис. 7). Такое расположение полезно при использовании цветовой кодировки значений метрик, ассоциированных с узлами графа, благодаря чему пользователь может получить представление о сложности выполняемой атаки. Древовидная компоновка графа удобна при изучении последовательности действий атакующего.



■ ***Рис. 6***. Представление результатов анализа уязвимостей в виде карты деревьев

■ Условные обозначения, используемые в системе оценки защищенности компьютерной сети

Обозначение	Описание
	Начальное положение злоумышленника
	Атомарное действие, имеющее разведывающий характер
	Сценарий, в котором не задействованы уязвимости
	Атакующее действие, использующее уязвимость



■ ***Рис. 7.*** Радиальное расположение узлов графа атак

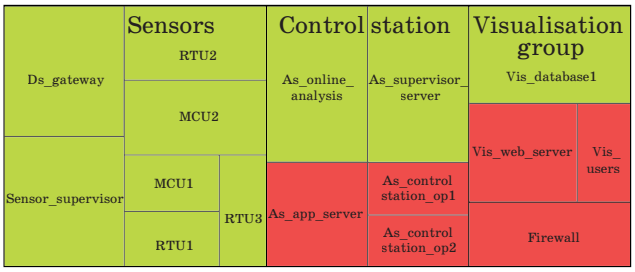
Семантическое масштабирование (агрегирование узлов графа). Для уменьшения сложности графа в системе используется механизм агрегации узлов графа, учитывающий его семантические и структурные свойства. Для формирования кластеров графа адаптирован подход, предложенный в работе [21]. В зависимости от таких свойств узла, как тип действия злоумышленника, хост, принадлежность группе/домену, вершины графа могут быть заменены на один мета-узел.

Используемые правила агрегации схематично представлены на рис. 8. Агрегирование узлов графа выполняется в интерактивном режиме. Пользователь имеет возможность определять степень агрегирования графа, задавая свойства и их последовательность применения для формирования мета-узлов.

Детали по требованию. При нажатии на узел графа мышью пользователь получает детальную информацию в отдельной вкладке окна. Эта информация включает тип атаки, хост, на котором выполняется атакующее действие, атакуемый хост, критичность хоста, описание уязвимости, вычисленные метрики безопасности (Ущерб, Уровень риска).

Подсветка и связывание. Данный визуальный эффект может быть использован для выделения пути в графе атак. При включении данного

■ ***Рис. 8.*** Иерархия правил агрегации узлов графа атак



■ ***Рис. 9.*** Представление скомпрометированных и защищенных узлов сети в виде карты деревьев

режима пользователь может выбрать узел графа, нажав на него указателем мышки, после чего все узлы, предшествующие и последующие выбранному, остаются цветными, а все остальные окрашиваются в оттенки серого.

Представление результатов моделирования атак в виде графов полезно при анализе последовательности действий злоумышленника, однако они не дают интуитивное представление о числе скомпрометированных узлов в сети. Для анализа достижимости узлов злоумышленником предлагается использовать карты деревьев, которые компактно представляют иерархическую структуру. Если, согласно результатам анализа защищенности компьютерной сети, узел может быть скомпрометирован, то соответствующий ему прямоугольник закрашивается красным цветом, в противном случае — зеленым.

На карте деревьев (рис. 9) размер прямоугольников соответствует уровню критичности узла для бизнес-процессов, а цветом обозначается состояние хоста. Поскольку при таком представлении пользователь не знает, какие уязвимости были использованы злоумышленником, он может получить данную информацию, нажав мышью на соответствующий прямоугольник карты деревьев.

Благодаря такому способу представления специалист может коррелировать выявленные уязвимости в компьютерной сети с числом потенциально скомпрометированных узлов, оценивая таким образом возможные последствия атаки.

Заключение

Анализ существующих программных решений по визуализации информации о защищенности компьютерной сети показал, что они предназначены для решения конкретной, достаточно узкой задачи, например, визуального анализа уязвимостей или моделирования атак. Для формирования полного понимания состояния защищенности системы пользователю необходимо применить нескольких таких инструментов, что может значительно усложнить работу системно-

го администратора и повлиять на ее эффективность в целом. Спроектированная авторами подсистема визуализации позволяет оценить уровень защищенности системы более полно и может быть использована как система поддержки принятия решения по планированию мероприятий по обеспечению безопасности, поскольку позволяет соотнести выявленные недостатки системы с возможными последствиями их эксплуатации и тем самым обоснованно определить наиболее критичные и требующие оперативного устранения уязвимости.

Работа выполняется при финансовой поддержке РФФИ, программы фундаментальных исследований ОНИТ РАН (проект № 2.2), государственного контракта 11.519.11.4008 и при частичной финансовой поддержке, осуществляемой в рамках проектов Евросоюза SecFutur и MASSIF.

Литература

1. **Keim D.** et al. Visual Analytics: Definition, Process, and Challenges // Information Visualisation, LNCS 4950. Springer-Verlag, 2008. P. 154–175.
2. **Новикова Е. С., Котенко И. В.** Механизмы визуализации в SIEM-системах // Системы высокой доступности. 2012. № 2. С. 91–99.
3. **Новикова Е. С., Котенко И. В.** Анализ механизмов визуализации для обеспечения защиты информации в компьютерных сетях // Тр. СПИИРАН. 2012. Вып. 4(23). С. 7–30.
4. **Котенко И. В., Саенко И. Б., Полубелова О. В., Чечулин А. А.** Применение технологии управления информацией и событиями безопасности для защиты информации в критически важных инфраструктурах // Тр. СПИИРАН. 2012. Вып. 1(20). С. 27–56.
5. **Чечулин А. А., Котенко И. В.** Комбинирование механизмов защиты от сканирования в компьютерных сетях // Информационно-управляющие системы. 2010. № 6(49). С. 21–27.
6. **Kotenko I., Chechulin A.** Common Framework for Attack Modeling and Security Evaluation in SIEM Systems // 2012 IEEE Intern. Conf. on Green Computing and Communications, Conf. on Internet of Things, and Conf. on Cyber, Physical and Social Computing, Besançon, France, Nov. 20–23, 2012 / Los Alamitos, California. IEEE Computer Society, 2012. P. 94–101.
7. **Bertini E., Hertzog P., Lalanne D.** SpiralView: Towards Security Policies Assessment through Visual Correlation of Network Resources with Evolution of Alarms // Proc. of the IEEE Symp. on Visual Analytics Science and Technology (VAST). 2007. P. 139–146.
8. **Foresti S.** et al. Visual Correlation of Network Alerts // IEEE Comput. Graph. Appl. 2006. Vol. 26. N 2. P. 48–59.
9. **Tran T., Al-Shaer E., Boutaba R.** PolicyVis: firewall security policy visualisation and inspection // Proc. of the 21st Conf. on Large Installation System Administration Conf. (LISA'07) / USENIX Association. Berkeley, CA, USA, 2007. P. 1–16.
10. **Mansmann F., Göbel T., Cheswick W.** Visual Analysis of Complex Firewall Configurations // Proc. of the 12th Intern. Workshop on Visualisation for Computer Security (VizSec'12). Seattle, WA, USA, 2012. P. 1–8.
11. **Marty R.** Applied Security Visualization. — N. Y.: Addison Wesley Professional, 2008. — 552 p.
12. **Montemayor J.** et al. Information Visualisation for Rule-based Resource Access Control // Proc. of Int. Symp. on Usable Privacy and Security (SOUPS), 2006. 2 p.
13. **Heitzmann A., Palazzi B., Papamanthou C., Tamassia R.** Effective Visualisation of File System Access-Control // Proc. of the 5th Intern. Workshop on Visualisation for Computer Security (VizSec'08). Berlin, Heidelberg: Springer-Verlag, 2008. P. 18–25.
14. **Harrison L.** et al. NV: Nessus Vulnerability Visualisation for the Web // Proc. of the 12th Intern. Workshop on Visualisation for Computer Security (VizSec'12). Seattle, WA, USA, 2012. P. 25–32.
15. **Nessus** vulnerability scanner website. http://www.tenable.com/ (дата обращения: 10.04.2013).
16. **Kotenko I., Stepashkin M.** Attack Graph based Evaluation of Network Security // Lecture Notes in Computer Science. 2006. Vol. 4237. P. 216–227.
17. **Noel S., Jacobs M., Kalapa P., Jajodia S.** Multiple Coordinated Views for Network Attack Graphs // Proc. of the IEEE Workshops on Visualization for Computer Security. IEEE Computer Society, 2005. P. 12.
18. **Williams L., Lippmann R., Ingols K.** An Interactive Attack Graph Cascade and Reachability Display // Proc. of the Workshop on Visualization for Computer Security, Sacramento, California, USA, 2007. Springer, Heidelberg. P. 221–236.
19. **Chu M.** et al. Visualizing Attack Graphs, Reachability, and Trust Relationships with NAVIGATOR // Proc. of the Seventh Intern. Symp. on Visualization for Cyber Security, Ontario, Canada. P. 22–33.
20. **Mell P., Scarfone K., Romanosky S.** A Complete Guide to the Common Vulnerability Scoring System Version 2.0 // Forum of Incident Response and Security Teams, June, 2007. P. 23.
21. **Homer J., Varikuti A., Ou X., McQueen M. A.** Improving Attack Graph Visualization Through Data Reduction and Attack Grouping // Proc. of the 5th Intern. Workshop on Visualisation for Computer Security (VizSec'08). Berlin, Heidelberg: Springer-Verlag, 2008. P. 68–79.

УДК 004.05

АНАЛИЗ ВРЕМЕННЫХ И СЛОЖНОСТНЫХ ХАРАКТЕРИСТИК ПАРОЛЬНОЙ АУТЕНТИФИКАЦИИ В ЗАЩИЩЕННЫХ ОПЕРАЦИОННЫХ СИСТЕМАХ СЕМЕЙСТВА Unix

Д. В. Юркин,
канд. техн. наук, доцент
Санкт-Петербургский государственный университет телекоммуникаций
А. В. Винель,
канд. техн. наук, ведущий научный сотрудник
ЗАО «НПФ ИНСЕТ», г. Москва
В. В. Таранин,
аспирант
Петербургский государственный университет путей сообщения

Описан подход к оценке вероятностно-временных характеристик протоколов аутентификации в операционных системах семейства Unix, основывающийся на теории вероятностных графов. Показано влияние действий нарушителя на работу протоколов аутентификации.

Ключевые слова *— криптографические протоколы, Unix OS, вероятностные графы.*

Введение

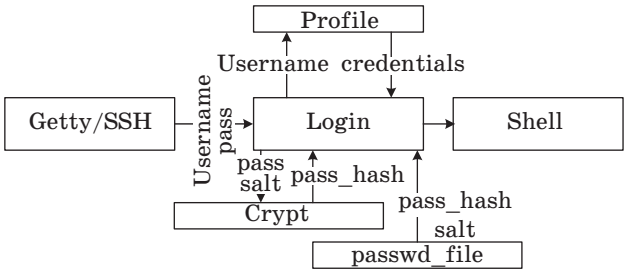
В мировой практике проектирования и построения защищенных информационных систем фактическим стандартом является использование Unix-подобных систем в качестве базовой операционной системы (ОС) для серверов и рабочих станций. Особый вклад в процесс эволюции защищенных ОС внесли ведущие разработчики и испытательные лаборатории систем обеспечения сетевой безопасности и средств защиты от несанкционированного доступа, которые на основании проводимых испытаний подтвердили отсутствие недекларированных возможностей, высокую отказоустойчивость встроенных механизмов защиты ОС. Портирование средств защиты информации Unix-подобных систем и широкий спектр поддерживаемых платформ привели к повсеместному использованию данных ОС производителями телекоммуникационного оборудования.

Правила реализации безопасной парольной политики и типовые настройки базовых встроенных механизмов управления доступом хорошо известны, однако вопрос анализа сложностных и временных характеристик успешного получения несанкционированного доступа к пользовательским и системным данным ОС на настоящий момент не подтверждены единым математическим доказательством.

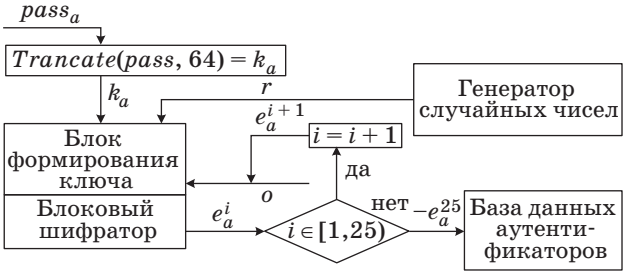
Описание способа парольной аутентификации

При предоставлении прав доступа к информационным ресурсам защищенной ОС возникает необходимость аутентификации пользователей для реализации механизмов дискретизации прав доступа. На данный момент наиболее распространенными и доступными ОС является семейство Unix, разработанное компанией Bell Laboratories. Встроенные механизмы зашиты таких ОС включают в свой состав протоколы парольной аутентификации [1], основой которых является верификация респондента по соответствию однонаправленного преобразования предъявленного пароля, приведенного в парольной таблице.

Взаимодействие программных модулей при аутентификации пользователей [2] осуществляется с вызова *getty* при непосредственном доступе с консоли и программных компонент пакета SSH при доступе с использованием сетевых средств управления программы *login*. Модуль *login* вызывается явно и замещает исходный интерпрета-



■ ***Рис. 1.*** Взаимодействие программных компонент при аутентификации субъектов в ОС



■ ***Рис. 2.*** Вычисление хеш-функции с использованием алгоритма DES

тор команд, после чего он выполняет проверку входных аутентификационных данных с использованием модуля криптографических преобразований *crypt* (или аналогичных). В случае успешной аутентификации *login* предоставляет доступ пользователю с соответствующими его профилю полномочиями к интерпретатору командной строки (рис. 1).

Вышеупомянутые криптографические преобразования могут быть реализованы выполнением функции шифрования с использованием ключа, полученного из пароля, конкатенированного с известной постоянной величиной и случайной последовательностью. В качестве однонаправленного преобразования может использоваться блоковое шифрование или ключевая хеш-функция. Ряд криптографических алгоритмов, реализуемых в схеме аутентификации (таблица), используются с добавлением к ключам случайных чисел.

В схеме однонаправленного преобразования по алгоритму DES (рис. 2) пароля $pass_a$ используется 25-кратное блоковое шифрование [3] нулевой последовательности o длиной 64 бита с добавлением битной случайной последовательности r с обратной связью, в качестве ключа k_a используются первые 64 бита пароля.

Добавление случайных чисел в алгоритм формирования ключей криптографического преобразования позволяет существенно затруднить атаку на базу аутентификаторов путем рандомизации его результатов. Поэтому алгоритмы перебора пароля при анализе базы данных увеличивают в общем случае трудоемкость вычислений в 2^r раза.

■ Используемые криптографические алгоритмы

Идентификатор алгоритма	Тип криптографического алгоритма
0	DES
1	MD5
2, $2a$, $2x$, $2y$	Blowfish
3	Алгоритм, совместимый с NT LAN Manager
4	SHA-1 (RFC 3174)
5	SHA-256 (RFC 4868)
6	SHA-512 (RFC 4868)

База аутентификаторов определяет соответствие идентификаторов пользователей, их символьных имен и соответствующих им хеш-функций паролей, а также другую информацию о пользователях и группах в системе. Этот массив данных представлен в виде текстовых файлов.

Существует два различных способа хранения паролей. Первый способ подразумевает общедоступное хранение аутентификаторов и хеш-функций паролей в едином файле вместе с реквизитами бюджетов пользователей. Второй, «теневой» способ ограничивает доступ пользователей к значениям хеш-функций паролей и определяет их размещение в отдельном файле, разрешенном на чтение и изменение только системным пользователям или процессам.

Очевидно, что «теневой» способ хранения значений однонаправленных криптографических функций паролей и случайных последовательностей позволяет увеличить защищенность системы аутентификации и повышает общий уровень робастности ОС относительно способов хранения аутентификационных данных, не использующих рандомизацию.

Однако при вышеописанном информационном обмене в процессе передачи данных инициатор провоцирует прямую компрометацию общего секрета. Это делает такую схему слабой аутентификации неприменимой в открытых каналах связи, а также предполагает ее использование только в доверенной среде передачи данных, что часто обеспечивается на практике посредством криптографической инкапсуляции передаваемых данных.

Методы формирования атак на протокол аутентификации

Рассмотрим применение методов теории вероятностных графов к моделированию различных схем взаимодействия участников информацион-

ного обмена в недоверенной среде передачи данных, состоящей из двух участников протокола (инициатора и респондента) и нарушителя. Предположим, что нарушитель имеет доступ к передаваемым сообщениям, поэтому может выполнять как перехват, так и подавление с подменой сообщения. Таким образом, взаимодействие корреспондентов информационного обмена происходит с участием посредника, который получает сообщения обоих легитимных сторон и может ретранслировать их без изменений, а может подменить любое сообщение на свое, и при этом факт подмены не будет замечен. Ориентированным графом покажем состояние схемы взаимодействия нарушителя и атакуемого легитимного корреспондента для протокола аутентификации (рис. 3).

Производящая функция перехода из состояния запроса аутентификации в состояние ее успешного завершения $f_r = 2^{-n}x_v^t$, а производящая функция перехода в начальное состояние протокола в случае предоставления неверных учетных данных равна $f_r^{err} = (1 - 2^{-n})x_v^t$. Общая производящая функция всего графа [4]

$$F(n) = \frac{f_r(n)}{1 - f_r^{err}(n)}.$$

Злоумышленник, получив запрос инициатора, пытается либо предугадать соответствующий ему ответ путем перебора общего секрета легитимных участников, либо просто угадать ответ. Предположим, что однонаправленное преобразование $R = f(S_{ab}, R)$ участники протокола выполняют идеально стойкой криптосистемой. Тогда вероятность того, что произвольно выбранное нарушителем значение S'_{ab} соответствует распределенному секрет, равна $P(S'_{ab} = S_{ab}) = 2^{-l(S_{ab})}$.

В случае, когда атакующий действует методом угадывания ожидаемого ответа, можно предположить, что все варианты отображения элементов множества запросов равновероятны. Тогда вероятность угадать ответ на i-й итерации $P(R'_i = R_i) = 2^{-l(R)}$. Таким образом, вероятность перехода из состояния r в состояние v будет $P(r \to v) = 2^{-n}$, где n — битовая длина перебираемой последовательности. Время, затрачиваемое верификатором на обработку одного запроса аутентификации, определяется величиной t_v. Согласно теории вероятностных графов [5], зависимость среднего времени успешного выполнения атаки угадыванием ответа от его длины с попыткой установления одной сессии протокола

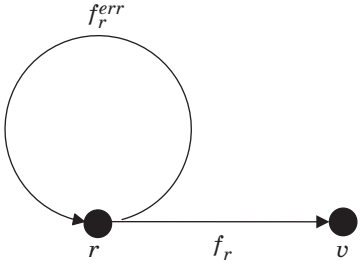
$$T_{e\_g}(n) = \frac{2^{-n} \cdot t_v \cdot (1 - (1 - 2^{-n})) + (1 - 2^{-n}) \cdot t_v \cdot 2^{-n}}{(1 - (1 - 2^{-n}))^2} = 2^n \cdot t_v.$$

Если атакующий действует методом перебора секретной последовательности, то очевидно, что с увеличением числа выполненных итераций протокола количество последовательностей, одна из которых является общим секретом легитимных корреспондентов, сокращается. Поэтому вероятность успешного перебора на i-й итерации $P(R'_i = R_i) = 2^{-(l(S_{ab})-i)}$. Таким образом, вероятность перехода из состояния r в состояние v равна $P(r \to v) = (2_s^n - i)^{-1}$, где n_s — битовая длина общего секрета. Зависимость среднего времени успешного выполнения атаки перебором секрета от его длины при попытке установления i сессий протокола

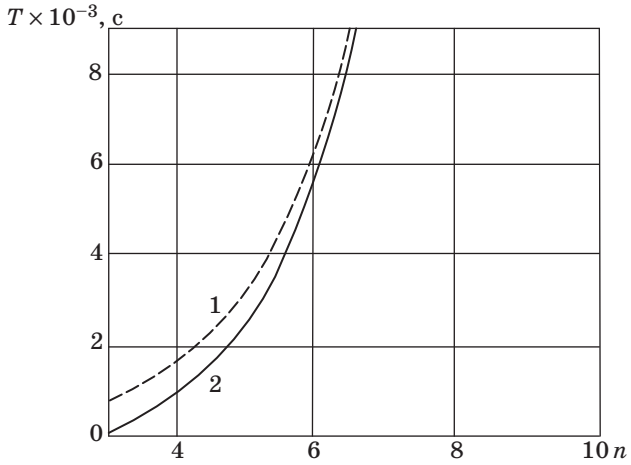
$$T_{e\_s}(n) = \frac{\frac{1}{2^n - i} \cdot t_v \cdot \left(1 - \left(1 - \frac{1}{2^n - i}\right)\right) + \left(1 - \frac{1}{2^n - i}\right) \cdot t_v \cdot \frac{1}{2^n - i}}{\left(1 - \left(1 - \frac{1}{2^n - i}\right)\right)^2} = (2^n - i) \cdot t_v.$$

Сравнение средних времен успешного выполнения атаки на протокол аутентификации перебором и угадыванием общего секрета представлено на рис. 4.

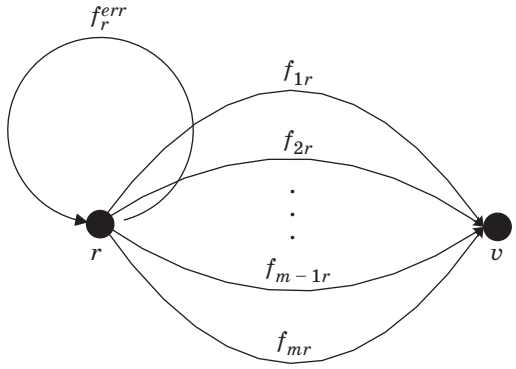
Однако наряду с последовательным выполнением итераций протокола атакующий также может одновременно начинать несколько сессий про-



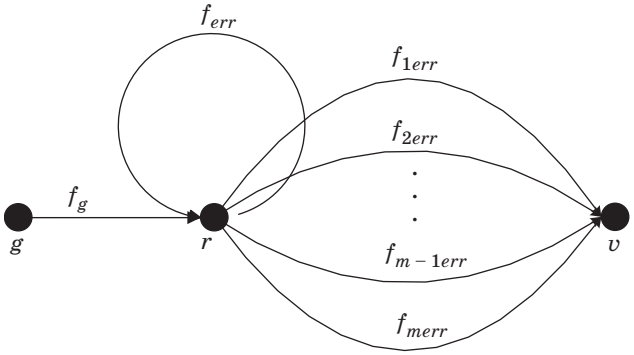
■ ***Рис. 3.*** Вероятностный граф протокола аутентификации



■ ***Рис. 4.*** Сравнение среднего времени атаки угадыванием (*1*) и перебором (*2*) общего секрета



■ ***Рис. 5.*** Вероятностный граф выполнения протокола аутентификации с инициализацией m параллельных сессий



■ ***Рис. 7.*** Вероятностный граф выполнения протокола аутентификации с инициализацией m параллельных сессий с предварительными вычислениями

токола, в этом случае вероятностный граф протокола будет иметь вид, показанный на рис. 5.

При этом общая производящая функция всего графа

$$F(n) = \frac{f_r(n)}{m(1 - f_r^{err}(n))},\ m \in \{1, \ldots, n\}.$$

Среднее время выполнения данного протокола для случая угадывания последовательности за одну (две, три и четыре) сессии $T_{e\_g}(n) = (2^n) \cdot t_v \cdot m^{-1}$ (рис. 6).

Если атакующий действует перебором общего секрета S_{ab} длиной $l(S_{ab}) = n$ с одновременным выполнением m сессий протокола, тогда ему на каждой попытке необходимо выполнить однонаправленное преобразование за время t_g, что, безусловно, увеличит время выполнения итерации протокола, которое станет равным $T_{e\_s}(n) = (2^n - m)(t_v + t_g) \cdot m^{-1}$. Создание дополнительной узловой точки формирования «словаря» возможных значений общего секрета перед вероятностным переходом означает вынесение детерминированной конструкции из циклической группы. Следовательно, не изменяя временной сложности итерации, можно понизить вычислительную сложность вероятностного цикла алгоритма. Вид вероятностного графа протокола показан на рис. 7.

В данном случае величина среднего времени

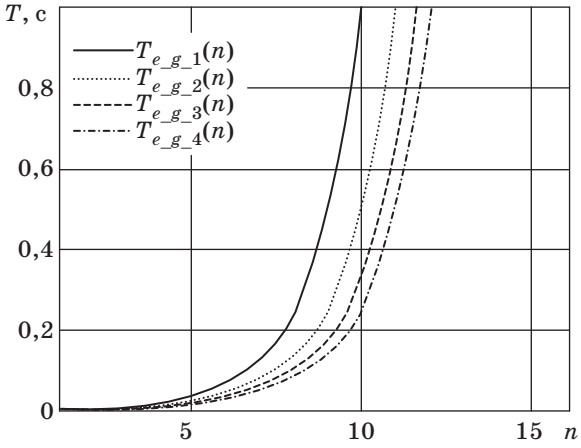
$$T_{e\_s}(n) =$$

$$= \frac{\frac{1}{2^n - m} \cdot (t_v + t_g)\left(1 - \left(1 - \frac{1}{2^n - m}\right)\right) + \left(1 - \frac{1}{2^n - m}\right) \cdot t_v \cdot \frac{1}{2^n - m}}{m\left(1 - \left(1 - \frac{1}{2^n - m}\right)\right)^2} =$$

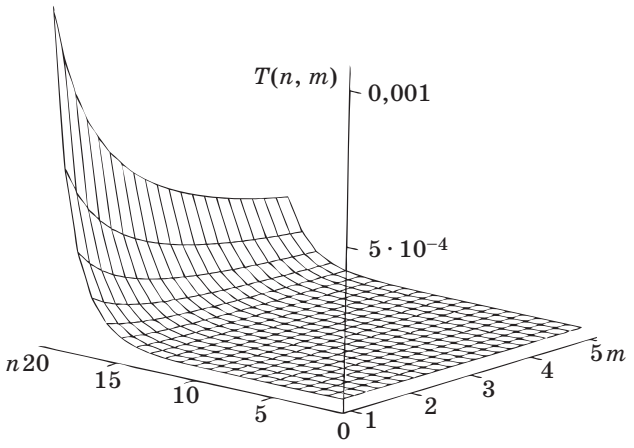
$$= m^{-1}((2^n - m) \cdot t_v + t_g).$$

Зависимость среднего времени выполнения протокола атаки от числа параллельных сессий и длины секрета в общем виде представлена на рис. 8.

Сравним результаты вероятностно-временного моделирования и теоретико-сложностных методов. Для этого произведем асимптотическую оценку функции трудоемкости алгоритма, определяющую сложность алгоритма и позволяю-



■ ***Рис. 6.*** Среднее время выполнения атаки на протокол угадыванием ответа



■ ***Рис. 8.*** Зависимость среднего времени выполнения атаки от числа сессий и длины секрета

щую выбрать предпочтения в использовании того или иного алгоритма для больших значений размерности исходных данных. Воспользуемся мажоритарной $O(g(n))$ оценкой, позволяющей дать верхнюю оценку для трудоемкости алгоритмов атак на протокол.

В случае атаки перебором ответа решаемая задача имеет экспоненциальную сложность, следовательно, асимптотическая оценка имеет вид $O(2^n)$.

Среднее время выполнения итерации протокола по алгоритму атаки, направленной на перебор ответа: $T(n) = \sum_{i=1}^{n} T(x) \cdot p(x)$, где $p(x)$ — вероятность появления входной последовательности x, а суммирование ведется по всем возможным входным последовательностям длины n. С учетом предположения о том, что при осуществлении однонаправленного криптографического преобразования все выходные последовательности равновероятны, $T(n) = T(x)$. Сопоставив значение среднего времени успешного выполнения протокола и асимптотическую ограничивающую сверху функцию $g(n)$, получим, что $T(n) = T(x)g(n)$, т. е. наиболее «близкой» (а в общем случае — равной) мажорирующей функцией будет сама трудоемкость алгоритма атаки с единичным постоянным множителем.

Таким образом, можно сделать вывод, что вероятностно-временной анализ посредством детального рассмотрения информационного взаимодействия корреспондентов позволяет находить методы линейного уменьшения среднего времени выполнения атаки перебором ответа, например запуском нескольких параллельных сессий.

Заключение

Результаты вероятностно-временного анализа алгоритмов атак на протокол парольной аутентификации в Unix-подобных системах соответствуют теоретико-сложностным оценкам трудоемкости выполнения этих атак, наглядно иллюстрируя возможные типы поведения нарушителя, дающие при этом четкое обоснование эффективности воздействия. Вероятностно-временные методы могут иметь широкое распространение при формировании свидетельства разработчика по стойкости функции безопасности объекта оценки AVA_SOF для проведения сертификационных испытаний средств защиты информации.

Литература

1. **Scott Mann, Ellen L. Mitchell, Mitchell Krell.** Linux system security. — Prentice Hall, 2003. — 617 p.
2. **Robert Morris, Ken Thompson.** Password Security: A Case History // Communications of the ACM. 1997. Vol. 22. P. 594–597.
3. **Philip Leong, Chris Tham.** Unix Password Encryption Considered Inside // USENIX. 1991. Vol. 3. P. 269–279.
4. **Nikitin V., Yurkin D., Chilamkurti N.** The influence of the cryptographic protocols on the quality of the radio transmission // Proc. of Intern. Conf. on Ultra Modern Telecommunications (ICUMT-2009), St.-Petersburg, Russia, Nov. 2009. P. 1–5.
5. **Никитин В. Н., Юркин Д. В.** Улучшение способов аутентификации для каналов связи с ошибками // Информационно-управляющие системы. 2010. № 6. С. 42–46.

УДК 681.3

ТИПЫ И ПРИЛОЖЕНИЯ ПРОТОКОЛОВ С НУЛЕВЫМ РАЗГЛАШЕНИЕМ СЕКРЕТА

А. А. Демьянчук,
научный сотрудник
А. Ю. Мирин,
канд. техн. наук, старший научный сотрудник
Н. А. Молдовян,
доктор техн. наук, заведующий лабораторией
Санкт-Петербургский институт информатики и автоматизации РАН

Представлены новые варианты протоколов с нулевым разглашением на основе трудности задач дискретного логарифмирования и факторизации. Обсуждается способ доказательства стойкости схем электронной цифровой подписи построением их путем преобразования протоколов с нулевым разглашением секрета. Предложен ряд новых протоколов с нулевым разглашением, включая двухпроходные.

Ключевые слова *— криптографический протокол, аутентификация, открытый ключ, электронная цифровая подпись, задача дискретного логарифмирования, задача факторизации.*

Введение

Протоколы с нулевым разглашением относятся к двухключевым криптосистемам и реализуют процедуры строгой аутентификации удаленных абонентов телекоммуникационных систем, что определяет области их практического применения для обеспечения информационной безопасности современных информационных технологий. Кроме того, протоколы данного типа могут быть использованы как базовый механизм построения алгоритмов электронной цифровой подписи (ЭЦП), а также для обоснования стойкости последних [1]. В типовом случае протоколы без разглашения секрета (с нулевым разглашением секрета) относятся к криптосхемам с открытым ключом (ОК) и представляют собой многораундовую процедуру, в которой типовой раунд выполняется за три интерактивных шага, из которых первые два включают использование случайных значений, а именно, раунд представляет собой выполнение следующих трех шагов:

1) доказывающий (субъект, подлинность которого доказывается в ходе протокола) генерирует разовый случайный секретный ключ, вычисляет по нему разовый ОК и направляет последний проверяющему (пользователю, желающему убедиться в подлинности доказывающего, т. е. в том факте, что доказывающий знает секретный ключ, связанный с ОК);

2) после получения разового ОК проверяющий генерирует запрос в виде случайного бита e и посылает его доказывающему;

3) в зависимости от значения полученного запроса доказывающий вычисляет ответ, который направляет проверяющему.

В каждом раунде доказывающий, если он является подлинным, дает правильный ответ с вероятностью, равной единице, а нарушитель может дать правильный ответ с вероятностью 0,5. Увеличивая число раундов в протоколе, можно понизить вероятность обмана до сколь угодно низкой величины. В случаях применения некоторых трудных задач за счет использования массива ОК (вместо одного ОК), принадлежащих доказывающему, удается построить одно раундовый протокол, включающий три шага, аналогичных указанным ранее. В таких случаях запрос проверяющего на втором шаге генерируется в виде битовой строки достаточно большого размера. Уменьшение числа раундов имеет существенное значение для практического применения протоколов аутентификации пользователей.

Важными для практического использования протоколов с нулевым разглашением является решение следующих задач: 1) уменьшение размера ОК в трехпроходных протоколах и 2) обеспечение достаточной очевидности отсутствия передачи информации о личном секретном ключе (ЛСК) в ходе выполнения протокола.

Термин «нулевое разглашение секрета» (речь идет о нулевой утечке информации о секретном ключе, связанном с ОК) следует понимать в том смысле, что данные, передаваемые доказывающим проверяющему, могли бы быть выработаны проверяющим самостоятельно. Следует отметить, что изначально объявляется некоторая утечка информации о ЛСК, по которому доказывающий вычислил свой ОК. Это состоит в том, что ОК является общедоступным, и по нему теоретически можно вычислить ЛСК, однако это практически нереализуемо, поэтому допускаемая утечка считается приемлемой. Трудность вычисления ЛСК по ОК является верхней границей стойкости протокола. Важным является недопущение какой-либо дополнительной утечки информации о ЛСК в ходе выполнения протокола. Если это обеспечивается, то многократное выполнение протокола практически не снижает его стойкости, т. е. наблюдение атакующим любого числа процедур выполнения протокола не упрощает ему задачу вычисления ЛСК по ОК (теоретически можно допустить, что случайно могут повториться случайные значения запроса, и тогда нарушитель может предоставить правильные ответы, которые он уже наблюдал ранее, однако вероятность такого повтора пренебрежимо мала).

В настоящей статье представлены подходы к построению двухпроходных протоколов с нулевым разглашением, обеспечивающие существенное сокращение размера ОК и очевидное доказательство нулевого разглашения секрета.

Итеративные протоколы

Хорошо известным и апробированным протоколом с нулевым разглашением секрета является протокол Фиата — Шамира [2], в котором доказывающий (пользователь, подтверждающий свою подлинность) доказывает, что он знает значение квадратного корня из некоторого числа t, которое служит ОК. Для того чтобы только подлинный владелец ОК знал значение корня из него, задача извлечения корня должна быть сложной. Это имеет место в случае, если корень извлекается по специально выбранному составному числу. Предполагается, что такое число формируется доверительным центром, который выбирает два больших простых числа p и q и вычисляет значение $n = pq$. Далее уничтожаются числа p и q, а число n используется для формирования пользователями своих ОК. Каждый пользователь выбирает случайное число s такое, что $1 \le s \le n - 1$, и вычисляет значение $t = s^2 \bmod n$. Протокол состоит из многократного повторения раунда, включающего следующие три шага:

1) доказывающий выбирает случайное число v такое, что $1 \le v \le n - 1$, вычисляет значение $q = v^2 \bmod n$, называемое фиксатором, и посылает его проверяющему;

2) проверяющий отправляет доказывающему случайный бит $r \in \{0, 1\}$;

3) доказывающий вычисляет значение $x = vs^r \bmod n$ и направляет его проверяющему. Проверяющий считает ответ положительным, если выполняется соотношение $x^2 = qt^r \bmod n$. В ходе осуществления протокола выполняются z шагов. Вероятность того, что нарушитель (который не знает секрета s) при выполнении одного раунда может дать положительный ответ, равна 2^{-1}, следовательно, вероятность того, что нарушитель может быть принят за пользователя, знающего секрет s, составляет 2^{-z}. Выбирая в протоколе достаточно большое число раундов проверки, можно сделать сколь угодно низкой вероятность обмана.

Для устранения необходимости наличия в протоколе доверительного центра можно предложить использовать трудность задачи извлечения корней большой простой степени по простому модулю со специальной структурой [3], а именно простое число p, имеющее структуру $p = Nk^2 + 1$, где разрядности чисел k и N равны, соответственно, $|k| \ge 160$ бит и $|N| \approx 864$ бит. Выбор таких значений разрядности связан с заданием минимально приемлемого уровня стойкости ЭЦП, равного 2^{80} операций модульного умножения [4]. В протоколе, основанном на трудности извлечения корней k-й степени по модулю p, каждый пользователь выбирает случайное число s такое, что $1 \le s \le p - 1$, и вычисляет значение своего ОК $t = s^k \bmod p$. Протокол состоит из z-кратного повторения следующего трехшагового раунда:

1) доказывающий выбирает случайное число v такое, что $1 \le v \le p - 1$, вычисляет значение $q = v^k \bmod p$ и посылает его проверяющему;

2) проверяющий отправляет доказывающему случайный бит $r = 1$ или $r = 0$;

3) доказывающий вычисляет значение $x = vs^r \bmod p$ и направляет его проверяющему.

Проверяющий считает ответ положительным, если выполняется соотношение $x^k = qt^r \bmod p$. Вероятность обмана составляет 2^{-z}. Следует отметить, что генерируемый на втором шаге случайный запрос проверяющего является принципиальным моментом протокола, поскольку при известном запросе потенциальный нарушитель может легко ввести в заблуждение проверяющего. Рассмотрим две возможные схемы действий нарушителя в одном раунде. В случае ожидаемого запроса $r = 0$ нарушитель выбирает произвольное число v и передает проверяющему значение $q = v^k \bmod p$. Если он получит от проверяющего запрос $r = 0$, то направляет правильный ответ

$x = v$. Однако правильно ответить на запрос $r = 1$ нарушитель не имеет возможности. В случае ожидаемого запроса $r = 1$ нарушитель выбирает произвольное число v и направляет проверяющему число $q' = v^k/t \bmod p$. Если он получит от проверяющего запрос $r = 1$, то направляет ответ $x' = v$, который будет принят проверяющим за правильный, поскольку

$$q't = (v^k/t)t = v^k = x'^k \pmod p.$$

Однако на запрос $r = 0$ нарушитель правильно ответить не сможет.

Трехпроходные протоколы

Рассмотрим реализацию трехпроходного протокола на основе итеративного протокола, использующего ОК вида $t = s^k \bmod p$, где простое 1024-битовое число $p = Nk^2 + 1$ при некотором простом k размером не менее 160 бит, и описанного в предыдущем разделе. В отличие от итеративного протокола в трехпроходном протоколе предполагается, что каждый пользователь в качестве своего ОК имеет h значений $t_i = s_i^k \bmod p$, где s_i — секретные значения, $i = 1, 2, ..., h$. Это позволяет объединить h однобитовых запросов итеративного протокола в единственный h-разрядный запрос E, что обеспечивает сокращение числа проходов до трех в следующем протоколе:

1) доказывающий выбирает случайное число v такое, что $1 \le v \le p - 1$, вычисляет значение фиксатора $q = v^k \bmod p$ и посылает его проверяющему;

2) проверяющий генерирует случайное h-разрядное число $E = (e_1, e_2, ..., e_h)$ и отправляет его доказывающему в качестве своего запроса;

3) доказывающий вычисляет значение

$$W = v\prod_{i=1}^{h} x_i^{e_i} \bmod p$$

и отправляет его в качестве своего ответа на полученный запрос.

Проверяющий считает ответ положительным, если выполняется соотношение

$$W^k = q\prod_{i=1}^{h} t_i^{e_i} \bmod p.$$

Вероятность обмана составляет 2^{-h}, что определяется следующими действиями нарушителя, пытающегося выдать себя за владельца ОК $(t_1, t_2, ..., t_h)$. Нарушитель генерирует случайный запрос $E' = (e'_1, e'_2, ..., e'_h)$ и случайный ответ W, после чего вычисляет значение фиксатора $q = W^k\prod_{i=1}^{h} t_i^{-e_i} \bmod p$. Затем он на первом шаге протокола отправляет проверяющему полученное значение фиксатора, ожидая получить запрос $E = E'$, что является событием, имеющим вероятность 2^{-h}. При наступлении такого события нарушитель успешно проходит процедуру аутентификации.

Двухпроходные протоколы

Рассмотрим двухпроходный протокол с нулевым разглашением, основанный на трудности задачи факторизации, для которого доказательство нулевой утечки секрета является достаточно очевидным. В качестве ОК доказывающего используется натуральное число n, равное произведению двух больших простых чисел r и q, составляющих его ЛСК. Идея доказательства состоит в том, что доказывающий передает проверяющему значение, которое вычислено последним до того, как оно было вычислено доказывающим, поэтому никакой новой информации от доказывающего не передается проверяющему. Протокол включает следующие два шага:

1) проверяющий генерирует случайное 36-битовое число k и, используя метод последовательного возведения в квадрат, вычисляет значение $T = 2^{2^k} \bmod n$, после чего передает доказывающему значение k в качестве своего запроса, на который он ожидает ответ доказывающего;

2) доказывающий выполняет две последовательные операции возведения в степень, в результате чего за короткое время вычисляет значения $K = 2^k \bmod L(n)$, где $L(n)$ — обобщенная функция Эйлера от числа n, и $T = 2^K \bmod n$. Затем он сразу направляет проверяющему значение T в качестве своего ответа на полученный запрос.

Если проверяющий получил правильное значение T, т. е. то значение, которое он вычислил до направления своего запроса доказывающему, в течение временного интервала, длительность которого не превышает некоторое пороговое значение Δ, то им делается вывод о подлинности доказывающего. Значение Δ выбирается достаточно малым. Его определяют исходя из того, что при выборе значений k размером от 30 до 36 бит время вычисления значения $T = 2^{2^k} \bmod n$ методом последовательного возведения в квадрат в 1000 и более раз должно превышать величину Δ. Применение данного протокола на практике требует учета возможных вычислительных ресурсов у потенциального нарушителя. Если предполагается возможность применения нарушителем специализированных производительных ЭВМ (применение многопроцессорных ЭВМ не дает эффекта, так как процесс последовательного возведения в квадрат не может быть распараллелен), то проверяющему требуется выбрать бóльшую разрядность для значения k. Это означает, что ему потребуется потратить больше времени на вычисление значения T. Выбираемая разрядность k

определяется также и быстродействием канала связи, используемого в протоколе. Чем больше быстродействие канала, тем меньшее значение Δ может быть выбрано, т. е. тем меньшая разрядность числа k может быть использована. Последнее означает уменьшение времени вычислений, выполняемых проверяющим до направления своего запроса доказывающему.

Обычно на практике один пользователь связывается со многими другими пользователями, подлинность которых он желает проверить. Это означает, что он должен для каждого из проверяемых установить свое пороговое значение Δ или взять общее пороговое значение $\Delta_{общ}$, равное максимальному времени, требуемому для получения ответа, по всем проверяемым пользователям. В первом случае требуется индивидуальная настройка параметров протокола аутентификации, но достигается меньшее среднее время вычислений на первом шаге. Во втором случае устраняется необходимость индивидуальной настройки параметров, но увеличивается время вычислений на первом шаге.

Использование технического параметра канала связи, связанного с его быстродействием, вносит существенные ограничения на области применения данного протокола. Его значение в выполненном исследовании состоит в том, что он показывает принципиальную возможность построения двухпроходных протоколов с нулевым разглашением и существенного сокращения размера используемого ОК (в десятки и сотни раз в зависимости от допустимого значения вероятности обмана). Устранение привязки к быстродействию канала достигается в следующих двух протоколах. При этом также обеспечивается элементарное доказательство того, что в ходе протокола не происходит утечка информации о секрете (доказательство нулевой утечки: проверяющий получает от доказывающего ответ на свой запрос, который он уже знает).

Первый протокол основан на схеме Диффи — Хеллмана открытого согласования общего секретного ключа [5] и описывается следующим образом. Как и в схеме Диффи — Хеллмана, системными параметрами протокола являются большое простое число p и соответствующий ему первообразный корень $\alpha < p$. Причем для обеспечения стойкости протокола размер числа p должен быть не менее 1024 бит, а разложение числа $p - 1$ на простые множители должно содержать, по крайней мере, один большой простой множитель длины не менее 160 бит. Открытые ключи пользователей генерируются следующим образом. Каждый пользователь выбирает случайный секретный ключ x (длиной не менее 160 бит) и вычисляет ОК y по формуле

$$y = \alpha^x \bmod p.$$

Задаваемая изначально утечка информации о секретном ключе заключается в том, что ОК делается общеизвестным, и любой желающий имеет принципиальную возможность однозначно вычислить значение секретного ключа x, хотя эта возможность практически нереализуема. Многократное выполнение протокола, приводимого далее, не уменьшает сложность реализации указанной потенциальной возможности больше чем на число операций, выполненных в процессе осуществления протокола. Протокол включает следующие два шага:

1) проверяющий генерирует случайное число k и вычисляет значения $U = \alpha^k \bmod p$ и $Z = y^k \bmod p$, где y — ОК доказывающего, после чего передает доказывающему значение U в качестве своего запроса, на который ожидает ответ доказывающего;

2) доказывающий вычисляет значение $Z = U^x \bmod p$, после чего направляет проверяющему значение Z в качестве своего ответа на полученный запрос.

Если проверяющий получил правильное значение Z, т. е. то значение, которое он вычислил до направления своего запроса доказывающему, то им делается вывод о подлинности доказывающего. Возможны различные варианты реализации аналогичных протоколов с использованием различных вариантов построении схемы Диффи — Хеллмана, например, основанных на трудности задачи дискретного логарифмирования в скрытой циклической подгруппе конечной некоммутативной группы [6, 7] или на трудности задачи дискретного логарифмирования на эллиптической кривой [8]. В последнем случае обеспечивается существенное уменьшение вычислительной сложности протокола.

Второй вариант реализации двухпроходного протокола с нулевым разглашением, свободный от привязки к временным параметрам канала связи, основан на использовании алгоритма открытого шифрования. Например, при использовании алгоритма открытого шифрования, подобного криптосистеме RSA [9], протокол описывается следующим образом. Личный секретный ключ и соответствующий ему ОК формируются пользователем следующим путем. Выбираются два больших, не равных между собой простых числа r и q, вычисляется произведение $n = rq$ и значение обобщенной функции Эйлера от $L(n)$, равной наименьшему общему кратному чисел $r - 1$ и $q - 1$. После этого генерируется случайное 32-битовое число e, взаимно простое с $L(n)$, и вычисляется число d, удовлетворяющее условию

$$ed \equiv 1 \bmod L(n).$$

Пара значений n и e является ОК. Тройка чисел r, q и d составляет ЛСК пользователя. Чи-

сла r и q должны иметь специальную структуру, в частности, они должны иметь разрядность не менее 512 бит, и каждое из чисел $r - 1$ и $q - 1$ должно содержать в своем разложении один большой простой множитель. Процедура открытого шифрования сообщения $M < n$ описывается формулой

$$C = M^e \bmod n.$$

Процедура расшифрования криптограммы C описывается формулой

$$M = C^d \bmod n.$$

Корректность процедуры расшифрования легко доказывается с использованием обобщенной теоремы Эйлера, согласно которой для любого числа M, взаимно простого с n, имеет место соотношение

$$M^{L(n)} \equiv 1 \bmod n.$$

Двухпроходный протокол с нулевым разглашением на основе данного алгоритма открытого шифрования имеет следующий вид:

1) проверяющий генерирует случайное сообщение $M < n$ и зашифровывает его по ОК (n, e) доказывающего: $C = M^e \bmod n$. Затем направляет доказывающему значение C в качестве своего запроса, на который ожидает ответ доказывающего;

2) доказывающий расшифровывает криптограмму C по своему ЛСК d, используя формулу $M = C^d \bmod n$. Полученное значение M доказывающий направляет проверяющему в качестве своего ответа на полученный запрос.

Если проверяющий получил правильное значение M, т. е. то значение, которое он сгенерировал на первом шаге протокола, т. е. до направления своего запроса доказывающему, то им делается вывод о подлинности доказывающего. Возможны различные варианты реализации аналогичных протоколов с использованием различных алгоритмов открытого шифрования, например алгоритмом открытого шифрования Эль-Гамаля [10]. При реализации протокола с нулевым разглашением на основе алгоритма Эль-Гамаля, построенного с использованием вычислений на эллиптической кривой, обеспечивается существенное повышение производительности протокола. Также могут быть построены производительные алгоритмы последнего типа при использовании алгоритмов открытого шифрования, основанных на трудности задачи дискретного логарифмирования в скрытой циклической подгруппе конечной некоммутативной группы [11].

Преобразование в схемы цифровой подписи

Протоколы с нулевым разглашением секрета имеют и другое важное применение, которое состоит в синтезе на их основе схем ЭЦП. Многие из протоколов с нулевым разглашением, а именно протоколы с предварительной передачей фиксатора проверяющему, могут быть легко преобразованы в схему ЭЦП, хотя в некоторых случаях размер подписи может быть настолько большим, что практическое применение протоколов будет нецелесообразным. Трехпроходные протоколы с нулевым разглашением позволяют получить схемы ЭЦП, пригодные для практического использования. Примером таких схем является протокол ЭЦП Фиата — Шамира [2, 12], выведенный из одноименного протокола с нулевым разглашением. Общая идея построения схемы ЭЦП на основе протокола с нулевым разглашением заключается в следующем. Подписывающий генерирует конкретное значение фиксатора. Далее, в зависимости от фиксатора и подписываемого документа, он вычисляет значение запроса, после чего вычисляет ответ на запрос. Пара чисел, включающая запрос и ответ, является цифровой подписью к документу. Таким образом, построенные алгоритмы ЭЦП относятся к рандомизированным криптосхемам, для которых к одному и тому же документу может быть выработано практически произвольное число различных подписей. Для того чтобы подделка подписи была практически невозможной, схема ЭЦП строится таким образом, что после вычисления значения запроса вычислительно трудно изменить значение фиксатора. Это может быть обеспечено вычислением значения запроса как значения стойкой хэш-функции от значения фиксатора с присоединенным к нему сообщением. В этом случае реализуется зависимость запроса от каждого бита фиксатора и каждого бита подписываемого документа. Преобразуем в соответствии с этой схемой трехпроходный протокол из предыдущего раздела в схему ЭЦП. Пусть подписывающий владеет ОК $(t_1, t_2, ..., t_h)$, где $t_i = s_i^k \bmod p$; k^2 — большой простой делитель числа $p - 1$; s_i — секретные значения, $i = 1, 2, ..., h$. Алгоритм формирования подписи к сообщению M зададим в следующем виде:

1) подписывающий генерирует случайное число $v < p$ и вычисляет значение фиксатора $q = v^k \bmod p$;

2) затем он, используя некоторую заранее оговоренную h-битовую хэш-функцию F_H, вычисляет значение $E = F_H(q, M) = (e_1, e_2, ..., e_h)$, являющееся первым элементом генерируемой ЭЦП;

3) далее он вычисляет значение

$$W = v\prod_{i=1}^{h} x_i^{e_i} \bmod p,$$

являющееся вторым элементом генерируемой ЭЦП.

Проверка подлинности ЭЦП (E, W) состоит в неявно заданной проверке выполнимости соотношения $W^k = q\prod_{i=1}^{h} t_i^{e_i} \bmod p$, используемого для проверки правильности ответа доказывающего в протоколе аутентификации с нулевым разглашением секрета. Поскольку значение q не задано в явном виде, процедура проверки подписи включает следующие шаги:

1) вычисляется значение фиксатора

$$q = W^k\prod_{i=1}^{h} t_i^{-e_i} \bmod p;$$

2) вычисляется значение хэш-функции $F_H(q, M) = E' = (e'_1, e'_2, ..., e'_h)$;

3) сравниваются значения E' и E. Если $e'_i = e_i$ для всех $i = 1, 2, ..., h$, то подпись принимается как подлинная. В противном случае подпись отклоняется как ложная.

Доказательство стойкости алгоритмов ЭЦП

Для обоснования безопасности схем ЭЦП, основанных на сложности задачи дискретного логарифмирования, может быть использован подход, основанный на синтезе протоколов с нулевым разглашением, из которых выводится схема ЭЦП, стойкость которой следует обосновать. В рамках данного подхода можно показать, что в ряде известных схем ЭЦП размер рандомизирующего параметра подписи может быть уменьшен в 2 раза без снижения стойкости. Эта возможность определяется тем, что для получения стойкой схемы ЭЦП достаточно получить низкую вероятность генерации ожидаемого запроса при фиксированном документе, а для устранения атак, связанных с возможностью модифицирования подписываемого документа в уравнение проверки ЭЦП, можно включить дополнительное значение другой хэш-функции (имеющей в 2 раза большую разрядность), вычисляемое только от документа. Рассмотрим обоснование стойкости схемы ЭЦП, построенной в предыдущем разделе.

В предположении, что изменение фиксатора после вычисления запроса $E = F_H(q, M)$ является практически невыполнимой задачей (это фактически является предположением о стойкости используемой хэш-функции), можно констатировать следующие факты.

1. Генерация произвольного числа подписей не упрощает задачу вычисления ЛСК подписывающего по его ОК.

2. Подделка подписи может быть выполнена с вероятностью 2^{-h} при использовании процедур с низкой трудоемкостью. Для получения большой вероятности удачной подделки подписи к заданному документу требуется выполнить порядка 2^h операций умножения по модулю p.

3. Для получения большой вероятности удачной подделки подписи в атаке с возможностью модифицирования подписываемого документа требуется выполнить порядка $2^{h/2}$ операций умножения по модулю p (это значение трудоемкости подделки определяется вычислительной сложностью нахождения коллизии хэш-функции с использованием парадокса о днях рождения [13]). Это означает, что для получения 80-битовой стойкости схемы ЭЦП требуется использовать, по крайней мере, 160-битовую хэш-функцию.

Эти факты позволяют говорить, что построенная из протокола схема ЭЦП является настолько стойкой, насколько стойким является протокол с нулевым разглашением, положенный в ее основу. Покажем, что разработанная схема ЭЦП может быть отнесена к доказуемо стойким криптосхемам в том смысле, что в предположении о стойкости используемой хэш-функции, т. е. о практической невозможности замены фиксатора после получения значения запроса E, алгоритм ее взлома имеет вычислительную сложность одного порядка со сложностью трудной задачи, положенной в ее основу. Пусть имеется некоторая атака, позволяющая подделать подпись без использования слабостей хэш-функции, т. е. вычислить значение ответа W по заданному значению запроса для различных используемых хэш-функций F_H и F'_H. Тогда, используя данную атаку, можно сгенерировать случайное значение фиксатора q и вычислить два правильных ответа W и W' для каждого из случаев использования F_H и F'_H, определяющих получение разных запросов $E = F_H(q, M)$ и $E' = F'_H(q, M)$. Правильные ответы удовлетворяют следующим соотношениям:

$$W^k = q\prod_{i=1}^{h} t_i^{e_i} \bmod p \text{ и } W'^k = q\prod_{i=1}^{h} t_i^{e'_i} \bmod p.$$

Выполнив деление первого соотношения на второе, получаем $\left(W/W'\right)^k = \prod_{i=1}^{h} t_i^{e_i - e'_i} \bmod p$.

Повторяя такую процедуру, можно получить достаточно большое число соотношений последнего вида, из которых легко найти представление элементов ОК ($t_1, t_2, ..., t_h$) в виде $t_i = s_i^k \bmod p$ для некоторого известного значения s_i (для всех значений $i = 1, 2, ..., h$). Это означает, что предположенная атака решает вычислительно трудную задачу извлечения корней большой простой степени k по модулю $p = Nk^2 + 1$, т. е. гипотетическая

атака имеет сложность одного порядка со сложностью решения использованной трудной задачи.

Отметим, что аналогичным способом можно дать формальное доказательство стойкости схем ЭЦП [4, 14] с малым размером ОК, которые основаны на трудности задачи извлечения корней большой простой степени по простому модулю со специальной структурой. Это можно выполнить, предложив протокол с нулевым разглашением, из которого затем вывести схему ЭЦП, стойкость которой требуется обосновать. Примеры такого обоснования стойкости других алгоритмов ЭЦП приводятся в работе [1].

Заключение

В настоящей статье предложен ряд протоколов с нулевым разглашением, которые представляют собой разные варианты решения поставленных исследовательских задач уменьшения размера ОК в протоколах с малым числом проходов и обеспечения достаточной очевидности отсутствия передачи информации о ЛСК в ходе выполнения протокола. Разработанные протоколы совмещают в себе использование ОК сравнительно малого размера и малое число проходов, что имеет существенное практическое значение. Также разработаны новые протоколы с нулевым разглашением, отличающиеся использованием трудности задачи извлечения корней большой простой степени по простому модулю, имеющему специальную структуру, и построена схема ЭЦП путем преобразования одного из предложенных протоколов. Дано формальное доказательство стойкости построенной схемы ЭЦП.

Литература

1. **Молдовян А. А., Молдовян Д. Н., Васильев И. Н., Головачев Д. А.** Протоколы с нулевым разглашением секрета и обоснование безопасности схем цифровой подписи // Вопросы защиты информации. 2011. № 4. С. 6–11.
2. **Молдовян Н. А., Молдовян А. А.** Введение в криптосистемы с открытым ключом. — СПб.: БХВ-Петербург, 2005. — 286 с.
3. **Молдовян Н. А.** Вычисление корней по простому модулю как криптографический примитив // Вестник СПбГУ. 2008. Сер. 10. Вып. 1. С. 101–106.
4. **Молдовян А. А., Молдовян Н. А.** Новые алгоритмы и протоколы для аутентификации информации в АСУ // Автоматика и телемеханика. 2008. № 7. С. 157–169.
5. **Diffie W., Hellman M. E.** New Directions in Cryptography // IEEE Transactions on Information Theory. 1976. Vol. IT-22. P. 644–654.
6. **Молдовян Д. Н.** Примитивы криптосистем с открытым ключом: конечные некоммутативные группы четырехмерных векторов // Информационно-управляющие системы. 2010. № 5. С. 43–50.
7. **Moldovyan D. N.** Non-Commutative Finite Groups as Primitive of Public-Key Cryptoschemes // Quasigroups and Related Systems. 2010. Vol. 18. P. 165–176.
8. **Болотов А. А., Гашков С. Б., Фролов А. Б.** Элементарное введение в эллиптическую криптографию: Протоколы криптографии на эллиптических кривых. — М.: КомКнига, 2006. — 274 с.
9. **Коутинхо С.** Введение в теорию чисел. Алгоритм RSA. — М.: Постмаркет, 2001. — 323 с.
10. **ElGamal T.** A public key cryptosystem and a signature scheme based on discrete logarithms // IEEE Transactions on Information Theory. 1985. Vol. IT-31. N 4. P. 469–472.
11. **Молдовян Д. Н.** Конечные некоммутативные группы как примитив криптосистем с открытым ключом // Информатизация и связь. 2010. № 1. С. 61–65.
12. **Fiat A., Shamir A.** How to prove yourself: Practical solutions to identification and signature problems // Advances in cryptology — CRYPTO'86. Springer-Verlag LNCS, 1987. Vol. 263. P. 186–194.
13. **Pieprzyk J., Hardjono Th., Seberry J.** Fundumentals of Computer Security. — Berlin: Springer-Verlag, 2003. — 677 p.
14. **Moldovyan N. A.** Digital Signature Scheme Based on a New Hard Problem // Computer Science J. of Moldova. 2008. Vol. 16. N 2(47). P. 163–182.

УДК 621.391.266

ФОРМИРОВАНИЕ И ОБРАБОТКА КОМПЛЕКСНОЗНАЧНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ В МНОГОКАНАЛЬНЫХ СИСТЕМАХ ПЕРЕДАЧИ ИНФОРМАЦИИ

Е. А. Григорьевых,
старший преподаватель, младший научный сотрудник
Р. Г. Хафизов,
доктор техн. наук, доцент
Поволжский государственный технологический университет, г. Йошкар-Ола

Предложен подход к организации многоканальной передачи информации с использованием комплекснозначных последовательностей. Показано, что применение комплекснозначных последовательностей, обладающих равномерным энергетическим спектром, позволяет устранить влияние соседних каналов. Рассмотрен принцип формирования кадра при многоканальной передаче информации на примере двоичных каналов.

Ключевые слова *— многоканальная связь, комплекснозначный сигнал, кодовое разделение каналов, пропускная способность.*

Введение

В многоканальных системах передачи информации, функционирующих по принципу кодового разделения каналов, используются сигналы, полоса которых во много раз превышает полосу частот при обычной передаче данных, например, в системах с частотным разделением каналов [1, 2]. Прием осуществляется оптимальным приемником, который для сигнала с полностью известными параметрами вычисляет корреляционный интеграл. Затем результат сравнивается с пороговым значением. Коррелятор (согласованный фильтр) производит «сжатие» спектра широкополосного входного сигнала путем умножения его на эталонную копию с последующей фильтрацией, что и приводит к улучшению отношения сигнал/шум на выходе коррелятора. Выбирая определенный ансамбль сигналов с «хорошими» автокорреляционными функциями (АКФ), можно обеспечить в процессе корреляционной обработки разделение сигналов.

В работе [3] показано, что большинство сигналов с идеальными АКФ и, соответственно, с равномерным энергетическим спектром являются комплекснозначными. Комплекснозначная последовательность $\mathbf{\Gamma} = \{\gamma(n)\}_{0,\ s-1}$ размерности s, состоящая из последовательности элементов $\gamma(n)$, т. е.

$$\mathbf{\Gamma} = \{\gamma(0), \gamma(1), \ldots, \gamma(s-1)\},$$

где $\gamma(n) = |\gamma(n)|\exp\{i\varphi(n)\}$, $n = 0, 1, \ldots, s-1$, будет обладать равномерным энергетическим спектром, т. е. $|\rho(0)|^2 = |\rho(1)|^2 = \ldots = |\rho(s-1)|^2$, в том случае, если составляющие его элементы равны

$$\gamma(n) = \frac{|\rho|}{s}\sum_{m=0}^{s-1}\left\{\cos\left[\frac{2\pi}{s}mn + \theta(m)\right] + i\sin\left[\frac{2\pi}{s}mn + \theta(m)\right]\right\},$$
$$n = 0, 1, \ldots, s-1.$$

Комплекснозначная последовательность, характеризующаяся равномерным энергетическим спектром, имеет только один значащий отсчет АКФ $\eta(0) = \|\mathbf{\Gamma}\|^2$, соответствующий главному лепестку функции. Здесь $\|\mathbf{\Gamma}\|^2$ — квадрат нормы последовательности. Все остальные отсчеты, образующие боковые лепестки АКФ, равны нулю. В данной работе рассмотрена возможность применения комплекснозначных последовательностей для организации многоканальной передачи информации.

Кодирование канальных сообщений комплекснозначными последовательностями

Использование комплекснозначных последовательностей с равномерным энергетическим спектром в системах передачи данных с кодовым разделением каналов основано на возможности идеального разделения каналов при корреляцион-

ной обработке сигнала за счет нулевой величины скалярного произведения между сигналами различных каналов, полученных путем циклического сдвига общей базовой кодирующей последовательности. Рассмотрим принцип формирования одного кадра при многоканальной передаче информации на примере двоичных каналов.

Применение в системах передачи данных с кодовым разделением каналов комплекснозначных последовательностей с равномерным энергетическим спектром основано на возможности идеального разделения каналов при корреляционной обработке сигнала за счет нулевой величины скалярного произведения между сигналами различных каналов, полученных путем циклического сдвига общей базовой кодирующей последовательности.

Имеется базовая кодирующая последовательность $\Gamma_б = \{\gamma(n)\}_{0,\ s\ -\ 1}$. Последовательности для нуля Γ_j^0 и единицы Γ_j^1 в j-м канале получаются циклическим сдвигом базовой последовательности соответственно на $2j$ и $2j + 1$ элементов, т. е.

$$\Gamma_j^0 = \left\{\gamma_j^0(n)\right\}_{0,\,s-1} = \left\{\gamma(n+2j)_s\right\}_{0,\,s-1};$$

$$\Gamma_j^1 = \left\{\gamma_j^1(n)\right\}_{0,\,s-1} = \left\{\gamma(n+2j+1)_s\right\}_{0,\,s-1},$$

где $(\bullet)_s$ — операция взятия по модулю s.

Групповой сигнал формируется путем векторного суммирования канальных сигналов, т. е.

$$\Gamma_{гр} = \left\{\gamma_{гр}(n)\right\}_{0,\,s-1} = \sum_{j=0}^{M-1}\Gamma_j^X = \left\{\sum_{j=0}^{M-1}\gamma_j^X(n)\right\}_{0,\,s-1},$$

где M — количество каналов; X — символ, передаваемый в j-м канале. На приемной стороне вычисляется скалярное произведение между эталонной кодирующей последовательностью каждого символа и принятым групповым сигналом $\Gamma_{гр}$:

$$\left(\Gamma_{гр}, \Gamma_j\right) = \eta = \sum_{n=0}^{s-1}\gamma_{гр}(n)\gamma_j^*(n).$$

Поскольку все используемые сигналы ортогональны, то взаимное влияние каналов при обнаружении и распознавании сигналов исключено. С учетом того, что кодирующие последовательности получены циклическим сдвигом базовой, операция обработки принятой кодовой последовательности сводится к вычислению ее циклической взаимной корреляционной функции (ВКФ) с базовой кодовой последовательностью:

$$\eta_r = \sum_{n=0}^{s-1}\Gamma_{гр}\gamma_0^*(n+r)_s,\ r = 0, 1, \ldots, 2M-1.$$

Рассмотрим принцип построения одного кадра многоканальной системы передачи информации с кодовым уплотнением каналов на примере комплекснозначной последовательности Γ = {1, 1, 1, 1, -0.5+0.866i, -0.5-0.866i, 1, -0.5-0.866i, -0.5+0.866i }. Каждому символу в каждом канале ставится в соответствие своя комплекснозначная последовательность, получаемая циклическим сдвигом базовой комплекснозначной последовательности на один элемент (таблица) [4].

При указанном методе кодирования количество каналов составляет $s/2 = 4{,}5$, где $s = 9$. Допустим, что в текущем кадре передается следующая комбинация {1, 1, 0, 1}, т. е. в первом канале передается 1, во втором — 1, в третьем — 0 и т. д.

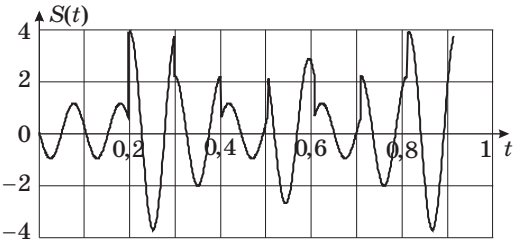
Групповой сигнал $\Gamma_{гр}$ = {-0.5+0.866i, -0.5+0.886i, 2.5-0.886i, 2.5-0.866i, 1, 1+1,732i, 1.25-0.866i, 2.5-0.866i, -0.5+0.866i, -0.5+0.866i, 2.5-0.866i, 2.5-0.866i, 1, 1+1.732i, 1}, элементы которого равны сумме элементов в соответствующих столбцах, задает последовательность, передаваемую в линию связи.

Комплекснозначные кодовые последовательности являются лишь математической моделью реального сигнала и не могут непосредственно использоваться в системах передачи и извлечения информации.

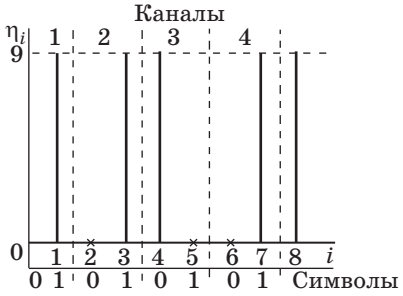
Реальным физическим носителем сигнала в пространстве может быть гармоническое колебание, какие-либо параметры которого изменяются по закону формирования элементов комплекснозначной последовательности [5]. В данной работе в качестве физического носителя был использован амплитудно-фазокодированный сигнал. При этом каждый элемент $\gamma(n)$, $n = 0, 1, \ldots, s - 1$, комплекснозначной последовательности $\Gamma = \{\gamma(n)\}_{0,\ s\ -\ 1}$ ассоциируется с n-м кодовым интервалом сигнала. В его пределах сигнал представляет собой отрезок синусоиды, амплитуда которой определяется модулем $|\gamma(n)|$, а начальная фаза — аргументом $\varphi(n)$ элемента $\gamma(n)$:

■ Таблица соответствия символов и комплекснозначных последовательностей в многоканальной системе связи

Канал	Символ	Кодовая комбинация
1	«0»	1, 1, 1, 1, -0.5+0.866i, -0.5-0.866, 1, -0.5-0.866i, -0.5+0.866i
	«1»	-0.5+0.866i, 1, 1, 1, 1, -0.5+0.866i, -0.5-0.866i, 1, -0.5-0.866i
2	«0»	-0.5-0.866i, -0.5+0.866i, 1, 1, 1, 1, -0.5+0.866i, -0.5-0.866i, 1
	«1»	1, -0.5-0.866i, -0.5+0.866i, 1, 1, 1, 1, -0.5+0.866i, -0.5-0.866i
3	«0»	-0.5-0.866i, 1, -0.5-0.866i, -0.5+0.866i, 1, 1, 1, 1, -0.5+0.866i
	«1»	-0.5+0.866i, -0.5-0.866i, 1, -0.5-0.866i, -0.5+0.866i, 1, 1, 1, 1
4	«0»	1, -0.5+0.866i, -0.5-0.866i, 1, -0.5-0.866i, -0.5+0.866i, 1, 1, 1
	«1»	1, 1, -0.5+0.866i, -0.5-0.866i, 1, -0.5-0.866i, -0.5+0.866i, 1, 1



■ ***Рис. 1.*** Пример фазокодированного сигнала суммарной комплекснозначной последовательности



■ ***Рис. 2.*** Результат вычисления ВКФ между базовой последовательностью и принятым сигналом

$$s(t)=\begin{cases}|\gamma(0)|\sin\left(\dfrac{2\pi t}{\tau_{\text{к.и}}}+\arg(\gamma(0)\right) \text{ при } 0\le t<\tau_{\text{к.и}};\\ |\gamma(1)|\sin\left(\dfrac{2\pi t}{\tau_{\text{к.и}}}+\arg(\gamma(1)\right) \text{ при } \tau_{\text{к.и}}\le t<2\tau_{\text{к.и}};\\ \ldots\\ |\gamma(s-1)|\sin\left(\dfrac{2\pi t}{\tau_{\text{к.и}}}+\arg(\gamma(s-1)\right)\\ \text{при } (s-1)\tau_{\text{к.и}}\le t<s\tau_{\text{к.и}},\end{cases}$$

где $\tau_{\text{к.и}}$ — длительность кодового интервала.

Пример амплитудно-фазокодированного сигнала, сформированного на базе суммарной комплекснозначной последовательности $\mathbf{\Gamma}_{\text{гр}}$, представлен на рис. 1.

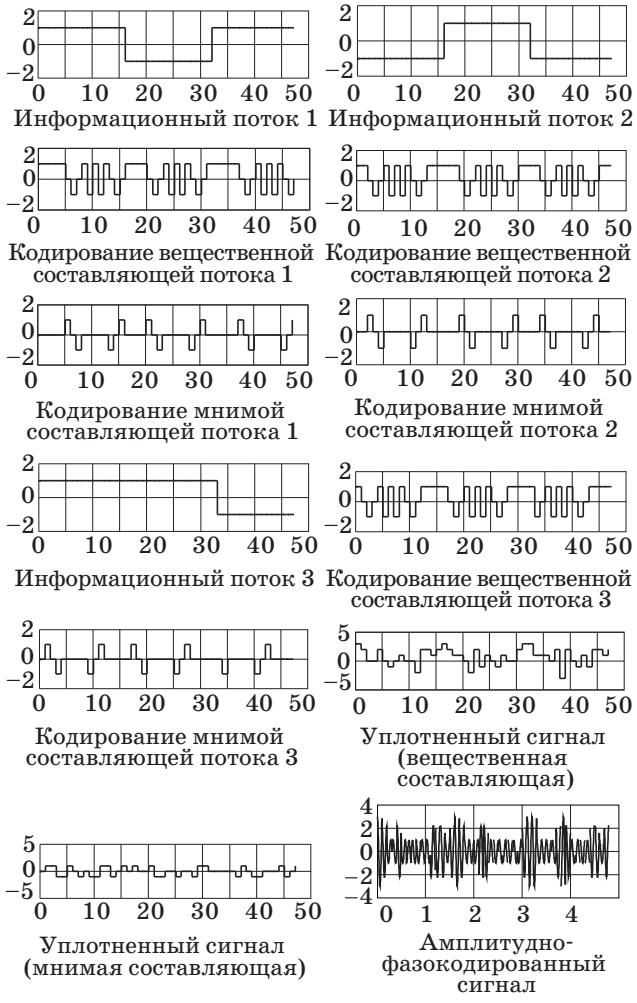
Для рассмотренного примера кодирования сигнал на выходе коррелятора будет иметь вид, показанный на рис. 2.

Положение ненулевых отсчетов на графике ВКФ соответствует сдвигам последовательностей относительно базовой.

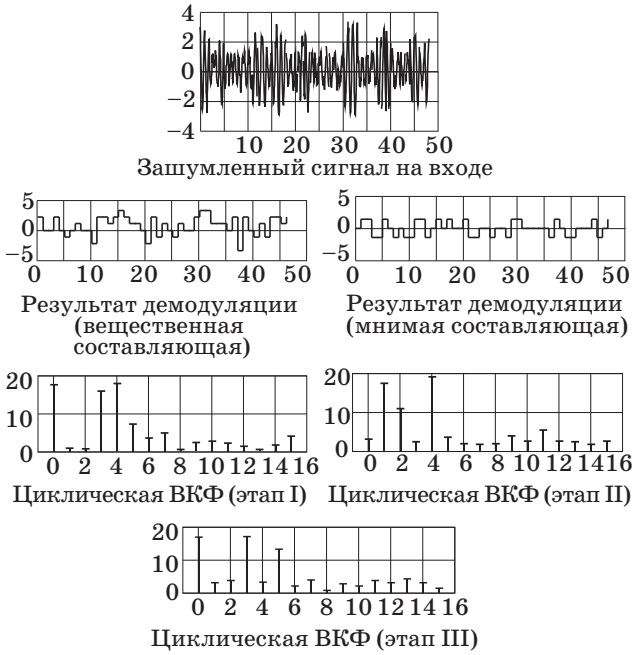
Таким образом, формирование группового сообщения сводится к суммированию отдельных кодирующих последовательностей, образуемых на основе базовой путем циклического сдвига и выбираемых в соответствии с передаваемым сообщением.

Организация многоканальной передачи информации с использованием комплекснозначных последовательностей

Рассмотрим передачу трех информационных потоков (рис. 3). Длительность информационных



■ ***Рис. 3.*** Уплотнение трех канальных сигналов



■ ***Рис. 4.*** Декодирование сложного сигнала

символов — 16. В качестве ортогональной функции взята комплекснозначная последовательность с равномерным энергетическим спектром размерности $s = 16$. Система ортогональных функций образована циклическим сдвигом элементов комплекснозначной последовательности. Каждому символу каждого потока соответствует своя последовательность.

Процесс демодуляции сложного сигнала в смеси с шумом представлен на рис. 4.

Как видно, корреляторы уверенно выделяют информацию даже на фоне помехи, несмотря на то, что передача ведется по разным каналам одновременно в общей полосе частот. Для эффективной работы системы критически важно отсутствие временных сдвигов между приходящим сложным сигналом и сигналом, выдаваемым генератором комплекснозначных последовательностей. Это достигается применением специальной синхронизирующей последовательности.

Анализ эффективности системы передачи данных

Предельная пропускная способность системы передачи с равномерной АЧХ и линейной ФЧХ в пределах полосы пропускания ΔF тракта передачи при наличии стационарного гауссова шума средней мощностью $P_{ш}$ и сигналов со средней мощностью $P_{с}$ определяется по формуле Шеннона

$$C = F\log_2\left(1 + P_{с}/P_{ш}\right).$$

При многоканальной передаче возникают специфические переходные помехи между каналами, обусловленные неидеальностью разделяющих устройств на приемной стороне и устройств формирования сигналов на передающей, линейными и нелинейными искажениями в групповом тракте передачи. Качество многоканальной системы с точки зрения переходных помех характеризуется величиной затухания $A_{ik} = 10\lg(P_i/P_{ik})$, где P_i и P_{ik} — мощности на входе влияющего и выходе подверженного влиянию каналов. Для мощности помех, наводимых i-м каналом на выходе k-го канала, имеем $P_{ik} = P_i 10^{-0{,}1A_{ik}}$, а общая мощность переходных помех $P_{п} = \mu P_{с}$, где $\mu = \sum_{i=1}^{N} 10^{-0{,}1A_{ik}}$ $(i \neq k)$ — коэффициент взаимных переходных помех между каналами. Если в формуле Шеннона учесть действия переходных помех, то

$$C = F\log_2\left(1 + \frac{P_{с}}{P_{ш} + \mu P_{с}}\right).$$

Поскольку обычно $P_{ш} \le \mu P_{с}$, то для пропускной способности системы многоканальной связи можно записать

$$C = F\log_2\left(1 + 1/\mu\right).$$

Последнее выражение позволяет учесть ограничение пропускной способности из-за действия переходных помех в отсутствии белого шума.

При использовании в качестве адресных сигналов дискретных комплекснозначных последовательностей, обладающих равномерным энергетическим спектром, величина затухания между влияющим i-м каналом и подверженным влиянию k-м каналом стремится к бесконечности. Коэффициент взаимных переходных помех μ при этом стремится к нулю.

Заключение

Применение комплекснозначных последовательностей с равномерным энергетическим спектром в многоканальных системах передачи информации теоретически дает возможность идеального разделения каналов при корреляционной обработке сигнала за счет нулевой величины скалярного произведения между сигналами различных каналов. Достоинством данной системы является тот факт, что взаимное влияние каналов при обнаружении и распознавании сигналов исключено, поскольку все используемые в системе сигналы ортогональны.

Литература

1. **Многоканальные** системы передачи / под ред. Н. Н. Баевой и В. Н. Гордиенко. — М.: Радио и связь, 1996. — 344 с.
2. **Цветков К. Ю., Коровин В. М., Косаревич Д. В.** Оптимальный ансамбль нелинейных сигналов для синхронных систем передачи информации с кодовым разделением абонентов // Информационно-управляющие системы. 2011. № 6(55). С. 40–44.
3. **Введение** в контурный анализ и его приложение к обработке изображений и сигналов / под ред. Я. А. Фурмана. — М.: Физматлит, 2002. — 592 с.
4. **Хафизов Р. Г., Григорьевых Е. А.** Применение комплекснозначных сигналов в системах асинхронной передачи данных // Телекоммуникации. 2007. № 10. С. 14–18.
5. **Хафизов Р. Г., Смирнов А. В., Григорьевых Е. А.** Исследование помехоустойчивости физических носителей комплекснозначных сигналов / МарГТУ, Йошкар-Ола, 2005. Деп. в ВИНИТИ 22.07.05. № 1070-В2005.

УДК 621.396

ПРИМЕНЕНИЕ МОДЕЛЕЙ СТРУКТУРНОЙ ДИНАМИКИ ПРИ РЕШЕНИИ ЗАДАЧИ РАСПРЕДЕЛЕНИЯ ЧАСТОТНО-ВРЕМЕННОГО РЕСУРСА СЕТИ СПУТНИКОВОЙ СВЯЗИ НА ОСНОВЕ СТАНДАРТА DVB-RCS

Е. А. Новиков,
канд. техн. наук, доцент
Военно-космическая академия им. А. Ф. Можайского, г. Санкт-Петербург

Рассмотрен стандарт использования спутникового ресурса DVB-RCS, в частности структура частотно-временного ресурса «обратных» каналов спутников-ретрансляторов. Определены основные недостатки используемых алгоритмов решения задачи оперативного распределения ресурса спутника-ретранслятора. Сформулирована и решена задача оптимального планирования ресурса «обратного» канала на основе моделей структурной динамики.

Ключевые слова *— спутниковая связь, DVB-RCS, MF-TDMA, обратный канал, частотно-временной ресурс.*

Введение

Инновационное развитие экономики России требует среди прочего интенсивного развития телекоммуникационной инфраструктуры, неотъемлемой частью которой в настоящее время является космическая составляющая. Использование спутниковых каналов передачи информации зачастую является единственным возможным вариантом организации связи и передачи данных. Анализ тенденций развития систем спутниковой связи говорит о росте функциональных возможностей и снижении стоимости спутникового оборудования как наземного, так и космического базирования. При этом у потребителя появляется возможность выбирать ту или иную технологию построения сети спутниковой связи исходя из своих потребностей и возможностей.

В настоящее время наиболее эффективным с точки зрения использования спутникового ресурса считается стандарт Digital Video Broadcasting — Return Channel via Satellite (DVB-RCS), определенный ETSI EN 301 790 v1.3.1, 2003 г. [1], а также его модификация DVB-RCS2 [2–4]. В России ведущим разработчиком бортового телекоммуникационного оборудования на основе стандарта DVB-RCS является ОАО «Научно-производственный центр «Вигстар». Основные принципы, заложенные в стандарте DVB-RCS, используются НПЦ «Вигстар» при разработке многофункциональных бортовых цифровых платформ для перспективных спутниковых платформ военного назначения.

Особенности решения задачи распределения частотно-временного ресурса «обратного» канала DVB-RCS

На этапе вхождения в связь спутниковый терминал (СТ) и центральная станция (ЦС) договариваются о применении тех или иных возможностей для обеспечения наибольшей функциональности. С технологической точки зрения стандарт DVB-RCS опирается на стандарт DVB-S2 при передаче информации в «прямом» канале и метод доступа с частотно-временным разделением каналов (MF-TDMA) при передаче информации в «обратном» канале, который лучше всего подходит для применения в спутниковых ретрансляторах с низкой линейностью и в условиях помех. Дополнительно в поток подмешивается управляющая информация и пакеты временной синхронизации. Обратный канал (в сторону ЦС) организуется с использованием MF-TDMA с инкапсуляцией MPE/DVB или LLC/ATM. Дополнительно оборудование ЦС осуществляет авторизацию терминалов, подстройку их параметров (уровня передачи, частоты, временных параметров).

Необходимо отметить, что стандарт [1] описывает физическую и логическую структуры кана-

ла спутниковой связи DVB-RCS, в то время как решение вопросов организации управления канальным ресурсом оставлено за рамками стандарта и, фактически, определяется каждым производителем оборудования самостоятельно. Очевидно, что от способа оперативного распределения канального ресурса спутника-ретранслятора зависят показатели качества обслуживания абонентского терминала. При этом специфика спутниковой связи, заключающаяся в возможном разрушении данных при передаче информации и централизованном распределении ресурса, определяет повышенные требования к уровню качества обслуживания.

Анализ работ, посвященных вопросам оперативного распределения ресурса спутника-ретранслятора [5–9] показал, что среди требований, предъявляемых к алгоритмам распределения ресурса, как правило, выделяются следующие:

1) время решения задачи распределения канального ресурса должно быть не более 100–200 мс [7];

2) алгоритм должен гарантировать выделение абоненту зарезервированный объем канального ресурса [5] с вероятностью не ниже заданной;

3) алгоритм должен гарантировать абоненту время задержки, не превышающее предварительно заказанного значения [5];

4) алгоритм должен гарантировать абоненту, что неиспользуемый ресурс канала будет распределен между абонентами пропорционально объему уже выделенного ресурса и что при подключении новых абонентов избыточный ресурс будет перераспределен [5].

При этом в работах [5, 10–17] рассматриваются эвристические алгоритмы распределения канального ресурса спутника-ретранслятора, являющиеся модификацией аналогичных алгоритмов для проводных сетей:

— иерархический циклический алгоритм (HRR) [11, 12];

— алгоритмы, основанные на фиксированном приоритете потоков [16];

— алгоритмы, основанные на растущем приоритете потоков [14].

Недостатками указанных алгоритмов являются отсутствие решения задачи распределения канального ресурса для многочастотного варианта, отсутствие строгого математического обоснования выбора и неоптимальность получаемых решений. Последнее обстоятельство говорит о неэффективном использовании имеющегося ресурса, особенно в условиях повышенного спроса абонентов на услуги спутниковой связи.

В работе [13] приводится постановка задачи линейного программирования, решение которой позволяет определить такие параметры фрейма, как количество частотных и временных позиций внутри фрейма, а также длительность позиции по частоте и времени.

Наиболее успешное решение задачи оптимального оперативного распределения ресурса спутника-ретранслятора приведено в работе [7]. При этом в ней изначально формулируется задача стохастического нелинейного целочисленного программирования, которая, однако, затем сводится к задаче нелинейной безусловной оптимизации с ослаблением ограничения целочисленности переменных до непрерывных значений и последующим эвристическим поиском целочисленных оптимальных значений. Полученное решение позволило обосновать возможный прирост коэффициента использования арендуемой пропускной способности на 16 %.

Характерными особенностями подходов [7, 8, 18] к решению задач оперативного распределения ресурса «обратного» канала являются, во-первых, предположение о случайном (стохастическом) характере поступления запросов абонентов сети спутниковой связи, а, во-вторых, использование для синтеза структуры фрейма методов только математического программирования, опирающегося на статические модели принятия решений.

Очевидно, что использование подхода, основанного на стохастическом детерминизме, для описания неопределенности системы и среды (граница, разделяющая исследуемую систему и среду, достаточно условна и определяется исследователем в каждом конкретном случае исходя из целей исследования), с одной стороны, прочно зарекомендовало себя как апробированный аппарат получения вероятностных оценок достижения целевых показателей системы, а, с другой стороны, нередко заставляет исследователей существенно упрощать реальные представления о системе и среде в угоду получению аналитических оценок. Тогда как реальное поведение системы и (или) среды зачастую трудно поддается формализации, а упрощенная формализация в итоге приводит к существенному снижению эффективности принятия решений в реальных системах.

В случаях, когда затруднительно пользоваться стохастическими методами ввиду отсутствия информации о распределении вероятностей параметров системы и среды, возможно применение нечетких методов [19] при математических расчетах в слабо формализуемых областях или детерминированных методов, вообще не использующих стохастическую или нечеткую информацию о системе и среде.

В свою очередь использование методов математического программирования существенно ограничивает возможности по выработке эффективных решений при организации оперативного управления ресурсом «обратного» канала, по-

скольку статические модели, по определению, не подчиняются принципу причинности и не обладают прогнозными свойствами. При этом динамические модели, например на основе дифференциальных уравнений, нашли широчайшее применение при решении достаточно сложных задач управления орбитальным движением космического аппарата, а также во многих других областях науки и техники.

Постановка и решение задачи оптимального планирования частотно-временного ресурса «обратного» канала

Предлагается в задачах распределения ресурса «обратного» канала по методу MF-TDMA использовать научно-методический аппарат на основе моделей структурной динамики [20], описываемых системой линейных дифференциальных уравнений в модифицированной форме Коши:

$$\dot{\mathbf{X}}(t) = \mathbf{B} \circ \mathbf{U}(t), \quad t \in [0, T_f], \tag{1}$$

где $\mathbf{X}(t) = \mathrm{col}(x_\nu(t), \nu = \overline{1, n})$ — n-мерный вектор объемов информации, передаваемых в рамках одного фрейма по каждому «обратному» каналу, $\nu = \overline{1, n}$ — номер СТ; $\mathbf{B} = \mathrm{col}(b_\nu, \nu = \overline{1, n})$ — n-мерный вектор скоростей передачи информации по «обратному» каналу, определяемых конфигурацией модема СТ (возможные значения компонентов вектора $\mathbf{B}$ приведены в таблице); $\mathbf{U}(t) = [u_{\nu\mu}(t) \in \in \{0, 1\}, \nu = \overline{1, n}, \mu = \overline{1, m}]$ — $n \times m$-мерная матрица управляемых параметров, определяющая конфигурацию обратного канала в момент времени $t \in [0, T_f]$, $\mu = \overline{1, m}$ — номер частотного интервала минимальной ширины; T_f — длительность фрейма; $[] \circ []$ — специальная операция матричного умножения, выполняемая по правилу $\dot{x}_\nu(t) = \sum_{\mu=1}^{m} b_\nu u_{\nu\mu}(t)$.

Примерные значения скоростей передачи информации приведены в таблице в зависимости от вида модуляции M и скорости кода R_k для фиксированных символьной скорости и минимальной ширины частотного интервала (например, для КА «Экспресс АМ-22» в Ku-диапазоне $\Delta f_\mu = 200$ кГц, $\mu = \overline{1, m}$, $R_s = 156$ ксимв/с, $T_f = 360$ с).

Модель распределения ресурса «обратного» канала (1) должна быть дополнена системой ограничений, к которым относятся технические и временные ограничения, а также краевые условия. Технические ограничения могут быть сформированы следующим образом:

— ограничение на использование одного частотного канала более чем одним СТ

$$\sum_{\nu=1}^{n} u_{\nu\mu}(t) - 1 = 0, \forall \mu = \overline{1, m}; \tag{2}$$

— ограничение на количество частотных каналов, занимаемых одним СТ:

$$\sum_{\mu=1}^{m} u_{\nu\mu}(t) - 1 = 0, \text{ или } \sum_{\mu=1}^{m} u_{\nu\mu}(t) - 2 = 0,$$
$$\text{или } \sum_{\mu=1}^{m} u_{\nu\mu}(t) - 4 = 0,$$
$$\text{или } \sum_{\mu=1}^{m} u_{\nu\mu}(t) - 8 = 0, \ \forall \nu = \overline{1, n}. \tag{3}$$

Временные ограничения могут быть сформированы следующим образом:

— интегральное временное ограничение

$$t \in [0, T_f]; \tag{4}$$

— дифференциальное временное ограничение, определяющее длительность интервала дискретизации, равного длительности одного тайм-слота (например, 4 мс):

$$\Delta t = t_{ts}. \tag{5}$$

Краевые условия должны быть заданы в форме «подвижного» правого конца:

$$x_\nu(T_f) \in \left[x_\nu^{\min}, x_\nu^{\max}\right], \nu = \overline{1, n}, \tag{6}$$

где $x_\nu^{\min} = b_\nu t_{ts}$ — объем информации, переданный в течение одного тайм-слота за время t_{ts}; $x_\nu^{\max}$ — объем информации, планируемый к передаче в соответствии с классом сервиса, заказанным абонентом, и включающий в себя объем информации, поставленный в очередь после распределения ресурса предыдущего фрейма.

Учет ограничений (2)–(6), накладываемых на модель (1), позволяет сформировать множество допустимых альтернатив управления U_Δ.

■ Скорость передачи информации в зависимости от конфигурации модема

Конфигурация модема	Вид модуляции								
	BPSK	QPSK				8PSK			
	Скорость кода R_k								
	1/2	1/2	2/3	3/4	7/8	1/2	2/3	3/4	7/8
Скорость передачи, Кбит/с	78	156	208	234	273	234	312	351	409,5

Показатель качества решения задачи распределения ресурса «обратного» канала может быть сформирован в виде функционала Майера [21]

$$J_1(T_f)=\sum_{i=1}^{s}\left(\alpha_i\left(\sum_{\nu\in Q_i}x_\nu(T_f)-y_i\right)^2\right), \qquad (7)$$

отражающего уровень требований к обеспечению дифференцированного качества QoS (Quality of Service) по классам сервиса, где $\mathbf{A}=\mathrm{col}(\alpha_i, i=\overline{1,s})$ — s-мерный вектор удельных весовых коэффициентов (штрафов) за отказ в полном обслуживании абонента соответствующего класса сервиса; $Q_i, i=\overline{1,s}$ — непересекающиеся множества индексов абонентов, сгруппированные по классам сервиса; $y_i=\sum_{\nu\in Q_i}x_\nu^{\max}, i=\overline{1,s}$ — суммарный объем информации, накопленный в буфере и сгруппированный по классам сервиса.

Следует отметить, что показатель качества (7) может быть преобразован к более удобному для дальнейшего использования виду

$$J_2(T_f)=\sum_{i=1}^{s}\left(\alpha_i\sum_{\nu\in Q_i}\left(x_\nu^{\max}-x_\nu(T_f)\right)^2\right). \qquad (8)$$

В таком виде показатель качества точно выражает штраф за невыполнение заявки на передачу определенного объема информации с учетом заказанного класса сервиса.

С учетом соотношения (8) может быть сформулирована задача динамической оптимизации

$$\mathbf{U}_{\mathrm{opt}}(t)=\arg\min_{\mathbf{U}\in U_\Delta}J_2(\mathbf{X},\mathbf{U},t). \qquad (9)$$

Динамическая интерпретация задачи оптимального планирования ресурса бортового ретранслятора позволяет для поиска решения сформулированной задачи управления использовать условие стационарности принципа максимума Л. С. Понтрягина [21]

$$\mathbf{U}_{\mathrm{opt}}(t)=\arg\max_{\mathbf{U}\in U_\Delta}H(\mathbf{P}(t),\mathbf{X}(t),\mathbf{U}(t)),\ \forall t\in T, \qquad (10)$$

где $H(t)=\sum_{\nu=1}^{n}p_\nu(t)\sum_{\mu=1}^{m}b_\nu u_{\nu\mu}(t)$ — функция Гамильтона; $p_\nu(t)$ — элементы s-мерного вектора сопряженных переменных $\mathbf{P}(t)$, совпадающего по размерности с вектором $\mathbf{X}(t)$.

Условие (10) дополняется каноническими соотношениями, позволяющими получить математические модели прямой и сопряженной систем:

$$\dot{\mathbf{X}}(t)=\left[\frac{\partial H}{\partial\mathbf{P}}\right]^{\mathrm{T}}=\mathbf{B}(t)\mathbf{U}(t);\quad \dot{\mathbf{P}}(t)=-\left[\frac{\partial H}{\partial\mathbf{X}}\right]^{\mathrm{T}}=\mathbf{0}. \qquad (11)$$

Для получения оптимального решения организуется итерационный процесс, предполагающий последовательное приближение от некоторого начального решения, называемого диспетчерским, к требуемому оптимальному. Известно [22], что скорость сходимости градиентных процедур поиска оптимальных решений, а именно к этому классу относятся процедуры решения краевых задач, во многом зависит от качества первого приближения, называемого диспетчерским решением, а точнее, от близости его к искомому оптимальному. В этих случаях часто пользуются так называемым условием трансверсальности

$$\mathbf{P}(T_f)=-\left[\frac{\partial h\left(\mathbf{X}(T_f),\mathbf{X}_f^{\max}\right)}{\partial\mathbf{X}(T_f)}\right]^{\mathrm{T}}, \qquad (12)$$

позволяющим определить состояние сопряженной системы в момент времени T_f, где функционал $h\left(\mathbf{X}(T_f),\mathbf{X}_f^{\max}\right)$ в нашем случае представлен показателем качества (8).

Хорошим опорным решением, в частности, может служить так называемое локально-оптимальное управление

$$\mathbf{U}_{l\,\mathrm{opt}}(t)=\arg\max_{\mathbf{U}\in U_\Delta}H(\mathbf{P}(t),\mathbf{X}(t),\mathbf{U}(t)),\forall t\in T, \qquad (13)$$

синтезируемое на основе условия стационарности (10) заменой вектора сопряженных переменных ее расчетным аналогом, полученным за счет модификации условий трансверсальности:

$$\mathbf{P}(t)=-\left[\frac{J_2(\mathbf{X},\mathbf{U},t)}{\partial\mathbf{X}(t)}\right]^{\mathrm{T}},\ \forall t\in T,$$

путем замены вектора $\mathbf{X}(T_f)$ на $\mathbf{X}(t)$, $\forall t\in T$. В результате такой замены получена однопроходная процедура синтеза локально-оптимального управления. При этом соотношение (13) является алгоритмом синтеза управления в форме обратной связи, которое выгодно отличается от оптимального программного управления, формируемого в результате решения краевой задачи, гибкостью и устойчивостью управляемого процесса.

Организация контура параметрической оптимизации процесса распределения частотно-временного ресурса «обратного» канала

Необходимо отметить, что открытым остается вопрос определения значений компонентов вектора удельных весовых коэффициентов $\mathbf{A}=\mathrm{col}(\alpha_i, i=\overline{1,s})$. При этом, как правило, подразумевается, что формирование вектора $\mathbf{A}$ происходит один раз на основе экспертной оценки степени важности того или иного класса сервиса. Очевидно, что такой подход может полностью нивелировать достоинства любого математического аппарата, примененного для решения поставленной задачи рас-

■ Организация контура параметрической оптимизации процесса распределения частотно-временного ресурса «обратного» канала

пределения ресурса «обратного» канала. В этой связи предлагается подойти к вопросу определения значений компонентов вектора **A** удельных весовых коэффициентов как к задаче параметрической оптимизации процесса распределения частотно-временного ресурса «обратного» канала (рисунок).

Для определения вектора удельных весовых коэффициентов могут применяться как детерминированные, так и нечеткие подходы. Механизм вычисления вектора удельных весовых коэффициентов должен учитывать динамику поступления запросов на обслуживание от абонентов сети. Другими словами, должен быть разработан алгоритм оценивания объема трафика, подлежащего передаче в составе следующего после планируемого фрейма. Для получения такого алгоритма могут применяться как детерминированные, так и нечеткие подходы. Среди детерминированных алгоритмов можно выделить алгоритм фильтр-дифференцирующего оператора [23], построенный на основе метода стохастической аппроксимации [24], обладающий минимальными потребностями в вычислительных ресурсах и нашедший применение в задачах обработки телеметрической информации. Использование нечетких подходов при решении задачи прогноза объема трафика, как правило, наталкивается на проблему численного решения дифференциальных уравнений, которая может быть решена на основе метода линеаризации [19].

Заключение

В статье проведен анализ особенностей решения задачи распределения частотно-временного ресурса «обратного» канала сети спутниковой связи на основе стандарта DVB-RCS. Сформулирована и решена задача оптимального планирования ресурса «обратного» канала на основе модели структурной динамики. Сформирована процедура локально-оптимального планирования, позволяющая получать решения, близкие к оптимальным, и не требующая организации итерационного вычислительного процесса. Сформулирована идея организации контура параметрической оптимизации, основанного на получении и учете прогнозных значений объемов трафика по классам сервиса. Решение задачи прогноза объема поступающего трафика предлагается осуществлять на основе детерминированных и нечетких подходов.

Литература

1. **ETSI EN 301 790.** Digital Video Broadcasting (DVB): Interaction channel for satellite distribution systems. 2000. http://etsi.org/deliver/etsi_en/301700_301799/01.05.01_60/en_301790v010501p.pdf (дата обращения: 20.12.2012).
2. **ETSI TS 101 545–1.** Digital Video Broadcasting (DVB), Second Generation DVB Interactive Satellite System (DVB-RCS2). Part 1: Overview and System Level specification. 2012. http://etsi.org/deliver/etsi_ts/101500_301599/01.01.01_60/ts_10154501v010101p.pdf (дата обращения: 20.12.2012).
3. **ETSI EN 101 545–2.** Digital Video Broadcasting (DVB). Second Generation DVB Interactive Satellite System (DVB-RCS2). Part 2: Lower layers for satellite standard. 2012. http://etsi.org/deliver/etsi_ts/101500_301599/01.01.01_60/ts_10154502v010101p.pdf (дата обращения: 20.12.2012).
4. **ETSI TS 101 545–3.** Digital Video Broadcasting (DVB), Second Generation DVB Interactive Satellite System (DVB-RCS2). Part 3: Higher layers for satellite specification. 2012. http://etsi.org/deliver/etsi_ts/101500_301599/01.01.01_60/ts_10154503v010101p.pdf (дата обращения: 20.12.2012).
5. **Березин К. Ю.** Гарантия качества обслуживания (QoS) в канале интерактивного взаимодействия цифрового спутникового телевидения // Электронный журнал «Исследовано в России». 2001. № 90. http://zhurnal.ape.relarn.ru/articles/2001/090.pdf (дата обращения: 31.01.2013).
6. **Генов А. А., Решетников В. Н.** Некоторые результаты имитационного моделирования мультисервисных бортовых цифровых платформ стандарта DVB-RCS // Программные продукты и системы / ЗАО НИИ «Центрпрограммсистем». Тверь, 2008. № 3. С. 38–41.

7. **Илюхин А. А.** Способ эффективного распределения ресурса пропускной способности спутниковых сетей интерактивного доступа // Вестник РГРТУ. 2009. Вып. 30. № 4. С. 16–21.
8. **Илюхин А. А., Дубровин А. Г.** Оптимизация структуры суперфрейма для запросных каналов в спутниковых сетях стандарта DVB-RCS // Изв. ОрелГТУ. 2009. № 2/52(563). С. 68–73.
9. **Марковский С. Г., Тюрликов А. М.** Использование идентификаторов абонентов для резервирования канала множественного доступа // Информационно-управляющие системы. 2008. № 2(33). С. 28–35.
10. **Gallager R. G., Parekh A. K.** A generalized processor sharing approach to flow control in integrated services networks: The single node case // ACM Transactions on Networking. 1993. N 1. P. 344–357.
11. **Hung A., Monpetit M.-J., Kesidis G.** ATM via Satellite: A Framework and Implementation // Wireless Networks. 1998. N 4. P. 141–153.
12. **Kalmanek C., Kanakita H., Keshav S.** Rate controlled servers for very high-speed networks // IEEE Global Telecommunications Conf. 1990. P. 300.31–300.39.
13. **Lee K.-D., Cho Y.-H., Lee S. J.** Optimal Design of superframe pattern for DVB-RCS return link // ETRI Jornal. 2002. Vol. 24. N 3. P. 251–254.
14. **Polyzos G. C., Sariowan H., Cruz R. L.** Scheduling for quality of service guaratees via service curves // Proc. of the Intern. Conf. on Computer Communications and Networks. 1995. N 3. P. 512–520.
15. **Stoica I., Zhang H., Ng T. S. E.** A Hierarchical Fair Service Curve Algorithm for Link-Sharing, Real-Time and Priority Service // Proc. of ACM SIGCOMM'97. 1997. P. 48–61.
16. **Verma D. C., Ferrary D.** A scheme for real time channel establishment in wide area networks // IEEE J. on Selected Areas in Communications. 1990. N 8. P. 368–379.
17. **Zhang L.** Virtual clock: A new traffic control algorithm for packet switched networks // ACM Transactions on Computer Systems. 1991. N 9. P. 101–124.
18. **Генов А. А., Решетников В. Н.** Адаптивное управление частотно-временным ресурсом космических аппаратов в сетях спутниковой связи // Информационные технологии и вычислительные системы. 2008. № 3. С. 55–62.
19. **Евдокимов А. В.** Численное решение нечетких дифференциальных уравнений методом линеаризации // Изв. Челябинского научного центра / Челябинский научный центр УрО РАН. Челябинск, 2003. № 4(21). С. 9–14.
20. **Соколов Б. В.** Комплексное планирование операций и управление структурами в АСУ подвижными объектами / МО РФ, 1992. — 232 с.
21. **Атанс М., Фалб П.** Оптимальное управление. — М.: Машиностроение, 1968. — 764 с.
22. **Моисеев Н. Н., Иванилов Ю. П., Столярова Е. М.** Методы оптимизации. — М.: Наука, 1978. — 352 с.
23. **Мануйлов Ю. С., Новиков Е. А., Павлов А. Р.** Решение задачи радиоконтроля орбиты космического аппарата на основе метода вероятностной аппроксимации // Радиотехнические и телекоммуникационные системы. 2012. № 1. С. 43–49.
24. **Батухтин В. Д., Майборода Л. А.** Разрывные экстремальные задачи. — СПб.: Гиппократ, 1995. — 362 с.

УДК 004.02:378

ПРИМЕНЕНИЕ ГЕНЕТИЧЕСКОГО АЛГОРИТМА ДЛЯ ОПТИМИЗАЦИИ УЧЕБНОГО ПЛАНА

О. Л. Курилова,
старший преподаватель
Ульяновский государственный университет

Представлен алгоритм оптимизации учебного плана в рамках компетентностного подхода. Разработан алгоритм построения матрицы смежности дисциплин, алгоритм ориентированного графа дисциплин, алгоритм нахождения самого длинного пути в графе. Продемонстрировано применение генетического алгоритма к многокритериальной задаче оптимизации учебного плана.

Ключевые слова *— компетенции, учебный план, ориентированный граф, генетический алгоритм, методы оптимизации.*

Введение

Компетентностный подход является на данный момент одним из основных направлений обновления содержания образования [1]. Понятие «компетенции» и их содержания разные исследователи определяют по-разному [1–5], но все они согласны с тем, что профессиональные компетенции в том или ином виде включают знания, умения, навыки и профессиональные качества личности.

Все множество изучаемых студентом дисциплин содержит в себе определенное количество компетенций, которые в явном виде не связаны между собой, а порядок изучения дисциплины (т. е. порядок усвоения компетенций) определяется кафедрами вуза при составлении рабочего учебного плана. Порядок следования дисциплин нередко назначают интуитивно, основываясь на опыте прошлых лет преподавания, и часто новые дисциплины без должной обработки занимают случайные места. Отследить связи между компетенциями, определенными Федеральным государственным образовательным стандартом высшего профессионального образования (ФГОС ВПО), без должного их анализа очень сложно ввиду индивидуальностей специалистов, формирующих содержание дисциплины и подготавливающих соответствующие учебники. В связи с изложенным разработка подхода к освоению компетенций на множестве изучаемых дисциплин с учетом их следования и «нарастания» объема связанных компетенций от начального до конечного этапов учебного процесса представляется важной и актуальной.

Взаимосвязи и формальное описание компетенций и дисциплин

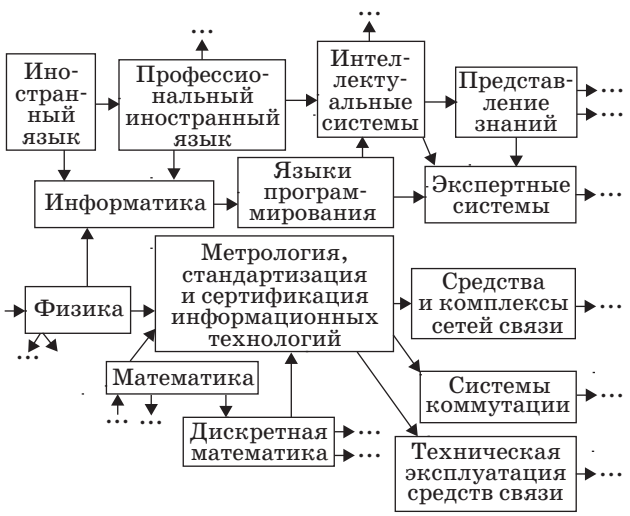
Взаимосвязь компетенций и дисциплин фиксируется в учебном плане и указывает на то, что составляющие компетенций являются продуктом изучения дисциплины. На рис. 1 представлена взаимосвязь элементов компетенций и дисциплин на примере дисциплины «Метрология, стандартизация и сертификация информационных технологий».

Взаимосвязь дисциплин прослеживается на основе входящих и исходящих компетенций. На рис. 2 показан фрагмент ориентированного графа, на котором для каждой дисциплины отображены предшествующие и следующие за ней дисциплины.

Требуется построить такой рабочий учебный план для формирования необходимых компетенций, который был бы эффективным и обеспечивал качественную подготовку выпускников вуза.

Исследование существующих учебных планов по различным специальностям показало, что распределение дисциплин по семестрам не является оптимальным и не всегда учитывает существующие взаимосвязи. Здесь под оптимальным учебным планом понимается такое распределение дисциплин по семестрам, в котором совокупность дисциплин удовлетворяет требованиям ФГОС ВПО, а именно определенному количеству часов, экзаменов, зачетов. При этом дисциплины должны быть связаны друг с другом элементами входящих и исходящих компетенций, причем нагруз-

■ ***Рис. 1.*** Взаимосвязь компетенций и дисциплин: ПК — профессиональная компетенция



■ ***Рис. 2.*** Фрагмент ориентированного графа взаимосвязи дисциплин на примере специальности «Информационные системы и технологии»

ка в семестрах должна быть распределена равномерно. Например, предметы «Основы теории управления», «Управление проектами», «Управление данными» изучаются одновременно во втором семестре второго курса, хотя логичнее было бы вначале изучить первый предмет, например в первом семестре, затем третий предмет во втором семестре, а второй предмет можно было бы изучить уже на третьем курсе.

Для оптимизации и построения учебных планов предлагается описание дисциплин и взаимосвязей между ними в следующем виде:

имеется множество дисциплин $D = \{D_1, D_2, ..., D_m\}$, каждый элемент множества D содержит подмножества компетенций: $D_j = \{K_1, K_2, ..., K_n\}$, $j = 1...m$, где $K_i = \{Th_i, Pr_i, C_i\}$ — компетенция i, $i = 1...n$, где Th_i — множество элементов теоретического знания; Pr_i — множество практических навыков; C_i — множество профессиональных качеств личности;

$D_j = \{Out_{Dj}, Inp_{Dj}\}$ — дисциплина, где $Out_{Dj} = \{Th_{Oj}, Pr_{Oj}\}$ — множество теоретических и практических элементов знания, формируемых в процессе изучения дисциплины; $Inp_{Dj} = \{Th_{Ij}, Pr_{Ij}\}$ — множество теоретических и практических элементов знания, необходимых для усвоения дисциплины.

Составляющими компетенций являются знания, умения и навыки $K_i = \{Th_i, Pr_i, C_i\}$. Например, для ПК-26, описанной в ФГОС ВПО [6], как готовность использовать математические методы обработки, анализа и синтеза результатов профессиональных исследований, характерны следующие компоненты:

ПК-26 = {знание основ теории вероятности и математической статистики; знание классических методов решения нелинейных методов уравнений; знание структуры погрешностей решения вычислительных задач; знание различных методов решения систем уравнений и т. д.}.

Каждая дисциплина D_i связана с дисциплиной D_j из множества D. Имеется множество связей $S = \{S_1, S_2\}$, через которые отыскиваются наиболее близкие друг другу дисциплины, где связь S_1 связывает дисциплины через знания, а S_2 — через умения и навыки.

Связь между D_j и K_i существует, если $Th_{Oj} \cap \cap Th_i \neq 0$, или $Th_{Ij} \cap Th_i \neq 0$, или $Pr_{Oj} \cap Pr_i \neq 0$, или $Pr_{Ij} \cap Pr_i \neq 0$.

Связь между D_j и D_y определяется следующим образом:

если $Inp_{Dy} \cap Out_{Dj} \neq 0$ и $Inp_{Dj} \cap Out_{Dy} = 0$, то D_j предшествует D_y (и наоборот);

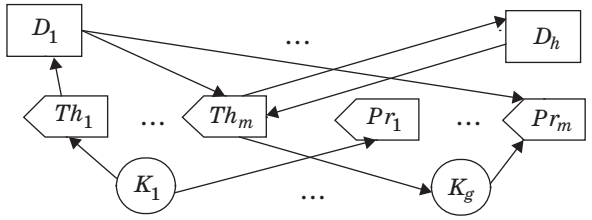
если $Inp_{Dy} \cap Out_{Dj} \neq 0$ и $Inp_{Dj} \cap Out_{Dy} \neq 0$, то D_j изучается в одном семестре с D_y;

если $Inp_{Dy} \cap Out_{Dj} = 0$ и $Inp_{Dj} \cap Out_{Dy} = 0$, то D_j не связаны с D_y.

Аналогично можно определить взаимосвязь компетенции и дисциплины.

Для более наглядного представления взаимосвязей дисциплин используем ориентированный граф.

Граф компетенций, элементов компетенций и дисциплин представлен на рис. 3. Связи в та-



■ ***Рис. 3.*** Граф взаимосвязи дисциплин и компетенций через элементы компетенций

ком графе формируются на основе изучения учебных планов и экспертных знаний преподавателей и специалистов.

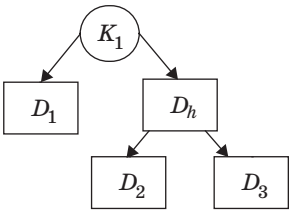
Если учитывать только связи между дисциплинами, то на рис. 4 показано, что граф преобразуется в дерево (без учета циклических связей между сильно связанными дисциплинами, которые изучаются параллельно), позволяющее понять, из каких дисциплин формируются отдельные компетенции.

Учебный план вуза для каждой специальности составляется на основе Типового положения об образовательном учреждении высшего профессионального образования [7] и ФГОС ВПО [6] для каждого направления подготовки. Например, для специальности 230400 «Информационные системы и технологии» (квалификация «бакалавр») используется ФГОС ВПО [6]: на стр. 2 указано количество часов и количество семестров, на стр. 7–14 определены названия дисциплин. Максимальное количество зачетов и экзаменов определено в п. 46 Положения [7].

Опишем формально эти требования. Каждую дисциплину можно представить в следующем виде:

$D_j = \{Name_j, Out_j, Inp_j, H_j, E_j, Z_j\}$, где $Name_j$ — имя дисциплины; Out_j, Inp_j — множества входящих и исходящих элементов знания; H_j — количество часов; E_j, Z_j — вид итогового контроля. Если $E(Dj) = 1$ и $Z(Dj) = 0$ — экзамен, $E(Dj) = 0$ и $Z(Dj) = 1$ — зачет. Тогда условия, накладываемые на дисциплины в семестре, выглядят следующим образом:

$$\sum_j^b H(D_j) \le R; \qquad (1)$$



■ ***Рис. 4.*** Дерево формирования компетенции через изучение дисциплин

$$\sum_j^b E(D_j) \le F; \qquad (2)$$

$$\sum_j^b Z(D_j) \le T; \qquad (3)$$

$$\sum_j^b \left(E(D_j) + Z(D_j)\right) \le P, \qquad (4)$$

где b — количество дисциплин, распределенных в семестре; R, F, T — соответственно количество часов, экзаменов и зачетов, допустимых в семестре; P — количество итоговых контрольных мероприятий, допустимых в семестре.

Алгоритм оптимизации учебного плана

К описанным дисциплинам можно применить методы оптимизации. Задача многокритериальной оптимизации заключается в том, что необходимо распределить последовательно дисциплины по семестрам с учетом определенных условий.

Существует целый класс оптимизационных методов. Условно все оптимизационные методы можно разделить на методы, использующие понятие производной (градиентные методы), стохастические методы, эвристические, эволюционные и прочие. При большом количестве параметров оптимизационные методы требуют больших временных ресурсов. Для задачи оптимизации учебного процесса пространство решений будет составлять $P = n!$, где P — мощность пространства поиска, а n — количество дисциплин. Поэтому предлагается использовать генетический алгоритм (ГА), который является одним из распространенных и наиболее употребляемых эволюционных методов.

Генетические алгоритмы — это процедуры поиска, основанные на механизмах естественного отбора и наследования. В них используется эволюционный принцип выживания наиболее приспособленных особей. Они отличаются от традиционных методов оптимизации несколькими базовыми элементами [8]:

— обрабатывают закодированную форму параметров задачи;

— осуществляют поиск решения, исходя из некоторого множества точек пространства возможных решений;

— используют целевую функцию;

— применяют вероятностные правила отбора.

В работе [9] показано применение ГА для оптимизации расписания, где в качестве критерия оптимизации использована система штрафов за недостатки в расписании групп и преподавателей.

В докладе [10] рассматривается последовательность дисциплин на основе ориентированно-

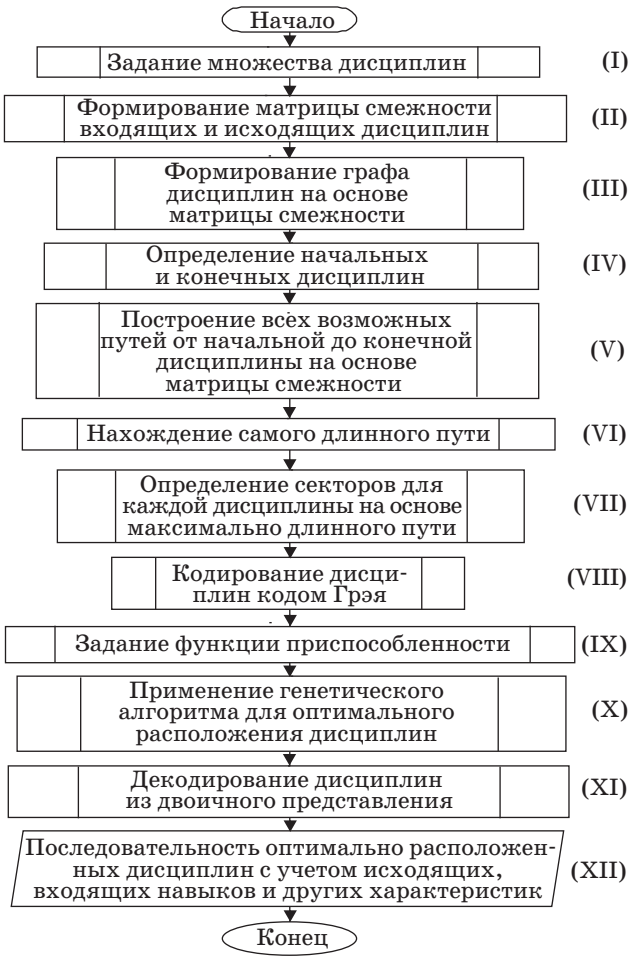
го графа, где определенным образом ранжируется каждая вершина, оценивается вес всего графа, и в результате целевая функция показывает относительное различие между оптимальным и текущим распределением дисциплин в учебном плане. Связь между дисциплинами определяется на основе экспертных заключений. В отличие от работ [9, 10], в данной статье описывается алгоритм автоматизированного построения ориентированного графа дисциплин на основе входящих и исходящих элементов компетенций, а также оптимизация учебного плана на основе целевой функции, удовлетворяющей определенным условиям (1)–(5).

Предлагаемый метод оптимизации учебного плана на основе формализованного компетентностного подхода можно представить в виде алгоритма. Суть метода состоит в формировании ориентированного графа, отображающего своими вершинами множество дисциплин и связей между ними (через множество компетенций), в котором отыскиваются самые длинные пути, включающие упорядоченный список дисциплин, которые обеспечивают преемственность компетенций от предыдущих дисциплин и передачу их последующим дисциплинам по цепочке пути (от начальных до конечным вершин графа). Далее путь на графе рассекается на фрагменты вершин в виде секторов, причем количество секторов совпадает с количеством семестров, а количество вершин в секторе совпадает с количеством дисциплин в семестре. Вводится процедура перераспределения дисциплин по всему графу. С этой целью используется целевая функция приспособленности конкретной особи (набора всех дисциплин) в популяции особей (наборе учебных планов), которая выражается удовлетворением атрибутов дисциплин требованиям (1)–(5). В качестве атрибутов дисциплин выступают количество часов, количество экзаменов, зачетов и принадлежность дисциплин определенному семестру.

Функцию приспособленности используем для оценки каждого учебного плана, после этого осуществляем сравнения значений этой функции и выбор максимального из них. Выбранный по максимальному значению функции приспособленности учебный план считается оптимальным. На рис. 5 представлен алгоритм, реализующий построение оптимального плана.

Опишем некоторые пункты этого алгоритма более подробно.

(II) Матрица смежности, назовем ее **A**, — это квадратная матрица размера $nd \times nd$ (nd — количество дисциплин), заполняется единицами и нулями по следующему алгоритму: если исходящий элемент компетенций для i-й дисциплины равен входящему элементу компетенций для j-й

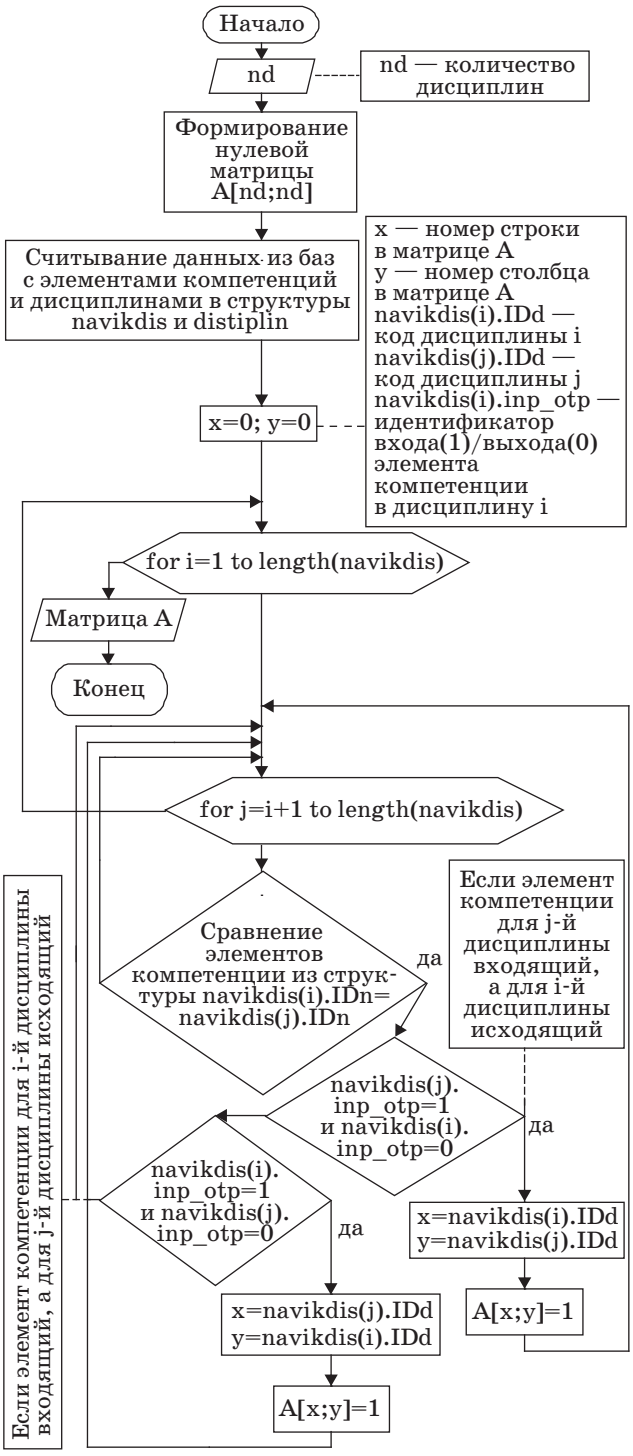


■ ***Рис. 5***. Алгоритм оптимизации учебного процесса

дисциплины, то $a[i, j] = 1$, в противном случае $a[i, j] = 0$. Алгоритм формирования матрицы смежности представлен на рис. 6. Матрица смежности необходима для формирования ориентированного графа дисциплин.

(III) Элементы матрицы смежности **A** являются вершинами ориентированного графа, который строится по следующему правилу: если $a[i, j] = 1$, то в ориентированном графе имеется ребро, соединяющее вершины i и j, поэтому i-я дисциплина является входящей для j-й дисциплины, т. е. дисциплина i должна быть изучена до дисциплины j; если $a[i, j] = 0$, то в ориентированном графе ребра нет.

(IV) При формировании учебного плана учитывается изначальный набор компетенций, с которыми учащийся пришел в вуз, например, те, которые получил в школе или колледже, поэтому при задании множества дисциплин на шаге (I) такие элементы компетенций будут помечены как начальные элементы компетенций. Считаем, что начальные дисциплины — это дисциплины, у которых количество входящих начальных эле-



■ ***Рис. 6.*** Алгоритм формирования матрицы смежности

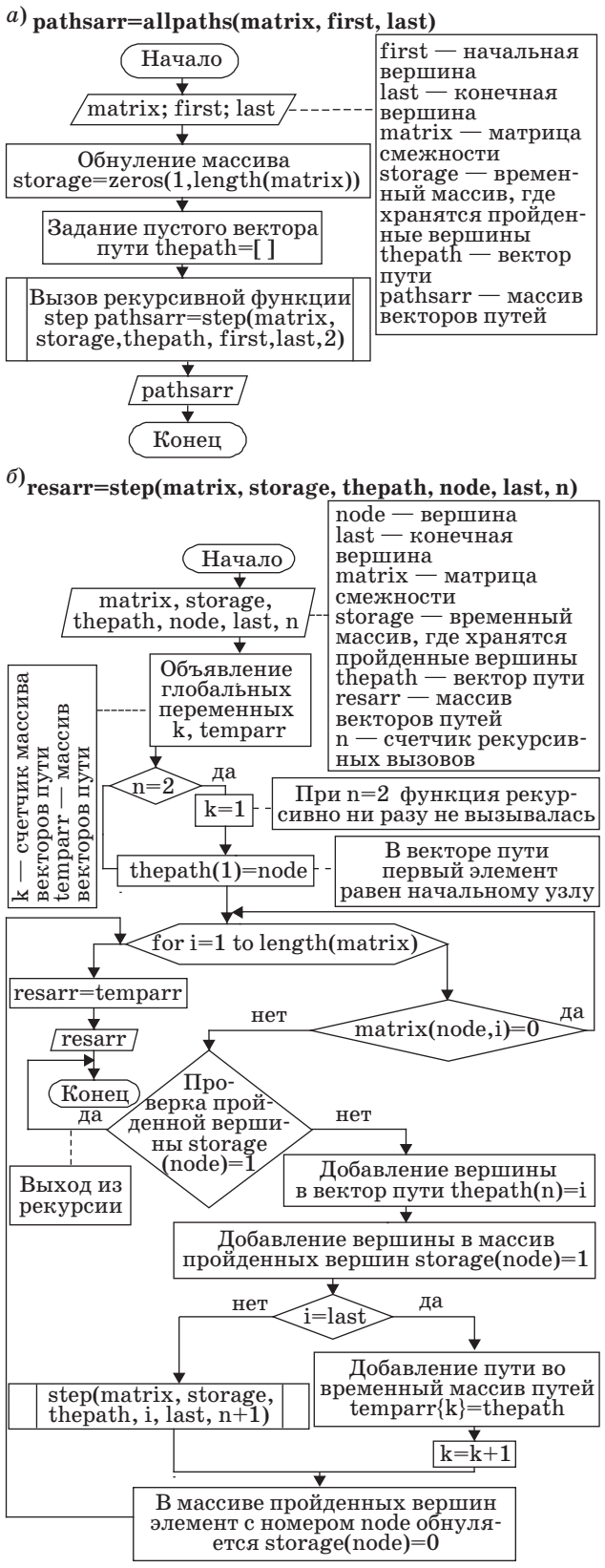
ментов компетенций минимально. Конечные дисциплины — это дисциплины, у которых количество исходящих элементов компетенций минимально. Признаком начальной дисциплины является минимальное количество единиц в матрице смежности в столбце, соответствующем данной дисциплине. Признаком конечной дисциплины является минимальное количество единиц в матрице смежности в строке, соответствующей данной дисциплине. Определение начальных и конечных дисциплин необходимо, чтобы между ними построить пути в ориентированном графе.

(V) Алгоритм построения представлен на рис. 7, *а*. Анализируются элементы матрицы смежности $a[i; j]$ и выбираются те, где $a[i; j] = 1$. Затем анализируется j-я строка и выбирается новый элемент в этой строке, где $a[j; k] = 1$. Процесс последовательной выборки единичных элементов образует путь. Далее по этой процедуре происходит полный перебор всех единичных элементов матрицы смежности и построение остальных путей с использованием рекурсии. Рекурсивный алгоритм представлен на рис. 7, *б*.

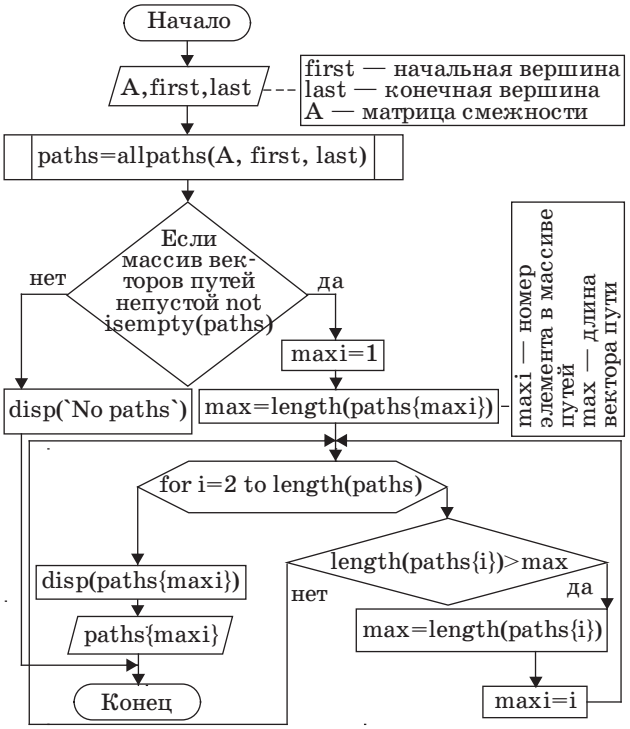
(VI) Из множества всех полученных путей выбирается путь с наибольшим количеством вершин. Алгоритм нахождения самого длинного пути в графе представлен на рис. 8. Самый длинный путь необходим для обеспечения полноты рекомендуемых стандартом дисциплин в учебном плане.

Существуют различные алгоритмы для работы с графами. Например, алгоритмы Джонсона и Дейкстры применяются для работы с ориентированным, взвешенным графом, где связи между вершинами закреплены постоянно и имеют вес, — эти алгоритмы находят кратчайшее расстояние в графе. В предлагаемых алгоритмах на основе матрицы смежности строится ориентированный, не взвешенный граф, а затем находится самый длинный путь. Элементы компетенций являются входящими и исходящими для дисциплин. Они могут изменяться, и именно на их основе строится матрица смежности, использующая в качестве исходных параметров наличие связей между дисциплинами. Поэтому предпочтение было отдано описанным выше алгоритмам.

(VII) Определение секторов для каждой дисциплины j на основе максимально длинного пути. Этот пункт необходим для предварительного распределения дисциплин по семестрам. Окончательное распределение дисциплин по семестрам происходит после применения ГА (X) и получения последовательности всех дисциплин. Поэтому полученный ориентированный граф надо разбить на секторы. Сектор — это интервал в пути графа, которому принадлежат дисциплины, причем каждый сектор имеет номер. Пусть длина самого длинного пути в графе (количество узлов в графе) L. Нумеруем каждый узел и для каждого узла длинной цепочки определяем сектор. Длина сектора вычисляется по формуле $\left\lceil \frac{L}{C} \right\rceil$. Например, если длина сектора 5, то узлы с 1-го по 5-й номер располагаются в первом секторе, а узлы с номе-



■ ***Рис. 7.*** Алгоритм нахождения всех путей в графе с использованием рекурсивной функции: *а* — между начальной и конечной вершинами; *б* — между двумя вершинами



■ ***Рис. 8.*** Алгоритм нахождения самого длинного пути в графе между двумя вершинами на основе матрицы смежности

ра 6 по 10 расположены во втором секторе и т. д. Узлы, не входящие в выбранную цепочку, могут располагаться в нескольких секторах. Каждому узлу графа надо сопоставить сектор i или интервал секторов $\left[S_{\min_j}; S_{\max_j}\right]$, где $S_{\min_j}$ — номер j-го узла с входной связью; $S_{\max_j}$ — номер j-го узла с выходной связью. К критериям оптимизации (1)–(4) добавляется еще один, который определяет номер сектора для каждой дисциплины:

$$\left\lceil \frac{S_{\min_j}+1}{\left\lceil \frac{L}{C} \right\rceil} \right\rceil \le i \le \left\lceil \frac{S_{\max_j}+1}{\left\lceil \frac{L}{C} \right\rceil} \right\rceil, \qquad (5)$$

где C — количество семестров; i — номер сектора.

(VIII) Прежде чем применять ГА, необходимо закодировать дисциплины. В классическом ГА применяется двоичное кодирование. Предлагается использовать код Грэя как наиболее простой и эффективный способ кодирования.

(IX) Функция приспособленности позволяет оценить степень приспособленности конкретной особи в популяции (конечное множество особей) и выбрать из них наиболее приспособленные в соответствии с эволюционным признаком выживаемости «сильнейших». При использовании ГА в задачах оптимизации функция приспособлен-

ности максимизируется, поэтому называется целевой функцией [8].

Критерием оптимального учебного плана служит целевая функция, выраженная посредством функции приспособленности:

$$F_{\text{ц}} = N_s \sum_{k=1}^{4} ball_k + \sum_{j=1}^{N_d \cdot N_s} ball_j \to \max,$$

где

$$ball_1 = \begin{cases} 1, \text{ если } \sum_{j}^{N_d} H(D_j) \le R \\ 0, \text{ если } \sum_{j}^{N_d} H(D_j) > R \end{cases}; \quad (6)$$

$$ball_2 = \begin{cases} 1, \text{ если } \sum_{j}^{N_d} E(D_j) \le F \\ 0, \text{ если } \sum_{j}^{N_d} E(D_j) > F \end{cases}; \quad (7)$$

$$ball_3 = \begin{cases} 1, \text{ если } \sum_{j}^{N_d} Z(D_j) \le T \\ 0, \text{ если } \sum_{j}^{N_d} Z(D_j) > T \end{cases}; \quad (8)$$

$$ball_4 = \begin{cases} 1, \text{ если } \sum_{j}^{N_d} (E(D_j) + Z(D_j)) \le P \\ 0, \text{ если } \sum_{j}^{N_d} (E(D_j) + Z(D_j)) > P \end{cases}; \quad (9)$$

$$ball_j = \begin{cases} 1, \text{ если } \left\lceil \frac{S_{\min_j} + 1}{\lceil L/C \rceil} \right\rceil \le i \le \left\lceil \frac{S_{\max_j} + 1}{\lceil L/C \rceil} \right\rceil \\ 0, \text{ если } i < \left\lceil \frac{S_{\min_j} + 1}{\lceil L/C \rceil} \right\rceil; \; i > \left\lceil \frac{S_{\max_j} + 1}{\lceil L/C \rceil} \right\rceil \end{cases}. \quad (10)$$

Единицей измерения целевой функции являются баллы, принимающие значения 1 или 0, причем значения $ball_1$, $ball_2$, $ball_3$, $ball_4$ накапливаются в целевой функции при соблюдении условий (6)–(9) для каждого семестра, а значения $ball_j$ — при соблюдении условий (10) для каждой дисциплины. Область значений целевой функции [0; *Ns**4 + *Nd***Ns*], где *Ns* — количество семестров, *Nd* — количество дисциплин в семестре. Целевая функция достигает максимума при выполнении условий (6)–(10), которым удовлетворяют дисциплины, входящие в учебный план (особь). Выходным параметром функции приспособленности является числовое значение, характерное для каждой особи (набора дисциплин).

Оптимальным считается учебный план, для которого достигнуто максимальное значение целевой функции.

Алгоритм функции приспособленности или целевой функции представлен на рис. 9.

В этом алгоритме учитываются многосеместровые дисциплины, т. е. дисциплины, количество часов которых превышает максимально допустимое число часов для дисциплин в семестре.

(X) В результате применения ГА некоторые дисциплины в учебном плане (особи) иногда повторяются, поэтому требуется устранить эту избыточность и заменить другими, которые не вошли в формируемый учебный план. Эту задачу можно решить во время работы ГА после операций скрещивания и мутации или после завершения работы ГА.

(XII) В результате применения данного метода формируется последовательность оптимально расположенных дисциплин с учетом исходящих и входящих элементов компетенций и других характеристик.

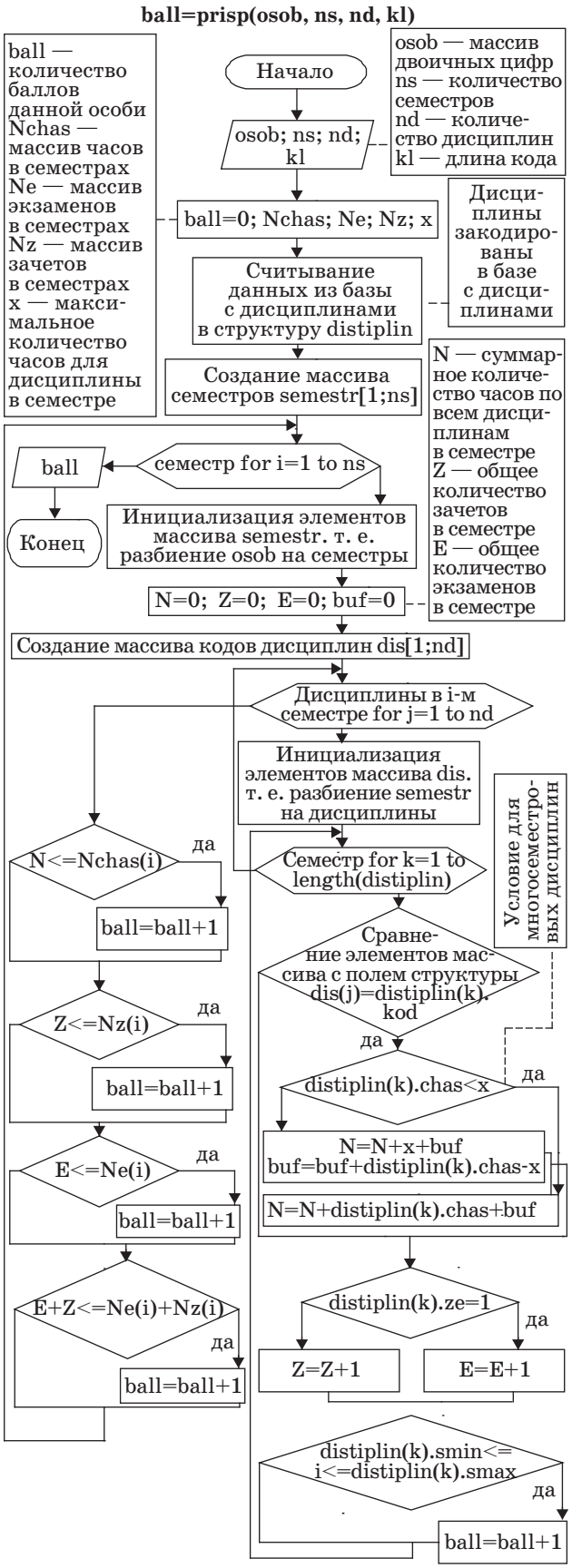
Апробация описанного метода была проведена с использованием программных инструментов MatLab [11, 12] для специальности 230400 «Информационные системы и технологии» (квалификация «бакалавр»).

При проведении экспериментов по оптимизации целевой функции были использованы следующие параметры: число дисциплин — 64, количество бит для кодирования одной дисциплины — 6 (поэтому число переменных в особи насчитывало 64*6 = 384), число семестров — 8, число особей — 60, число поколений — 100.

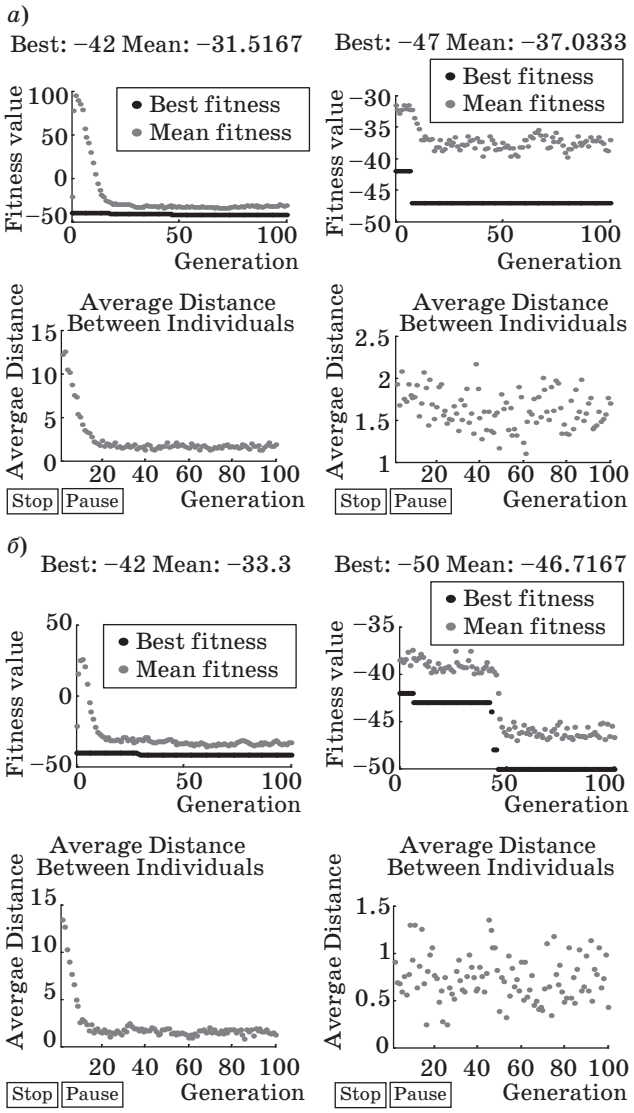
В *эксперименте 1* использовались следующие операторы ГА: равномерный кроссовер с вероятностью 0,8, одноточечная мутация, пропорциональный отбор. На рис. 10, *а*, слева, показаны наилучшие значение функции и расстояние между популяциями. Так как модуль по применению ГА *Genetic Algorithm and Direct Search Toolbox* в MatLab решает только задачи минимизации, для нахождения максимума целевой функции используется результат функции приспособленности, только со знаком минус.

Поскольку ГА является стахостическим, т. е. производятся случайные выборки, в результате применения ГА каждый раз получаются разные результаты и строится новая популяция, но чтобы достичь лучших результатов, можно в качестве исходной популяции использовать результаты для конечной популяции из предыдущих расчетов. После проведения этой операции значение целевой функции увеличилось (рис. 10, *а*, справа).

Далее был проведен *эксперимент 2*. Использовались следующие операторы ГА: двухточечный



■ ***Рис. 9***. Алгоритм функции приспособленности



■ ***Рис. 10***. Результаты применения ГА после эксперимента 1 (*а*) и 2 (*б*)

кроссовер, одноточечная мутация, турнирный отбор. Значение целевой функции снова увеличилось (рис. 10, *б*).

После проведения ряда экспериментов был получен оптимальный учебный план.

Заключение

Предлагаемый метод позволяет:

— выявить несогласованность в последовательности формирования дисциплин в учебном плане;

— сформировать последовательность с учетом исходящих и входящих элементов компетенций и других характеристик;

— визуально представить все связи между дисциплинами на основе ориентированного графа дисциплин;

— оценить готовые учебные программы на полноту охвата компетенций;

— автоматизированно построить оптимальный учебный план.

Этот метод оптимизации учебного процесса рекомендуется применять в высших, средних, среднеспециальных учебных учреждениях, на курсах подготовки и переподготовки специалистов.

Литература

1. **О Концепции** модернизации российского образования на период до 2010 года: Приказ № 393 от 11.02.02 / МО Российской Федерации. http://elementy.ru/Library9/pr393.htm (дата обращения: 25.02.2013).
2. **Лайл М. Спенсер, Сайн М. Спенсер.** Компетенции at work. Модели максимальной эффективности работы. — М.: HIPPO, 2005. — 372 с.
3. **Уиддет С., Холфорд С.** Руководство по компетенциям. — М.: ГИППО, 2008. — 228 с.
4. **Морозова Г. Б.** Психологическое сопровождение организации и персонала. — М.: Речь, 2006. — 400 с.
5. **Курилова О. Л., Смагин А. А., Липатова С. В.** Методы оценки компетенций выпускника вуза // Уч. зап. Ульяновского государственного университета. Сер. Математика и информационные технологии. 2012. Вып. 1(4). С. 246–257.
6. **Федеральный** государственный образовательный стандарт высшего профессионального образования по направлению подготовки 230400 «Информационные системы и технологии» (квалификация (степень) «бакалавр»). http://www.edu.ru/db-mon/mo/Data/d_10/prm25-1.pdf (дата обращения: 25.02.2013).
7. **Типовое** положение об образовательном учреждении высшего профессионального образования (высшем учебном заведении): утв. постановлением Правительства РФ от 14 февраля 2008 г. № 71. http://www.fgosvpo.ru/uploadfiles/npo/20110419090913.pdf (дата обращения: 25.02.2013).
8. **Рутковская Д., Пилиньский М., Рутковский Л.** Нейронные сети, генетические алгоритмы и нечеткие системы. — М.: Горячая линия-Телеком, 2006. — 454 с.
9. **Яндыбаева Н. В.** Генетический алгоритм в задаче оптимизации учебного расписания вуза // Современные наукоемкие технологии. 2009. № 11 С. 97–98. http://www.rae.ru/snt/?section=content&op=show_article&article_id=5657 (дата обращения: 06.03.2013).
10. **Дроздов Н. А.** Оптимизация учебных планов // Информационные технологии в образовании (ИТО-ЧЕРНОЗЕМЬЕ): Междунар. науч.-практ. конф., 2008 г. http://ito.edu.ru/2008/Kursk/V/V-0-5.html (дата обращения: 06.03.2013).
11. **Кривилев А.** Основы компьютерной математики с использованием системы MATLAB. — М.: Лекс-Книга, 2005. — 485 с.
12. **Панченко Т. В.** Генетические алгоритмы. — Астрахань: Астраханский университет, 2007. — 87 с.

УДК 004.932.2

О ВЕКТОРНОМ КВАНТОВАНИИ ИЗОБРАЖЕНИЙ

Е. А. Крук,
доктор техн. наук, профессор
М. Б. Сергеев,
доктор техн. наук, профессор
Санкт-Петербургский государственный университет аэрокосмического приборостроения

Рассматриваются возможности использования кодов, исправляющих ошибки, в задачах векторного квантования для сжатия изображений. Предлагаются методы выбора кода и процедуры проведения преобразований изображения, позволяющие согласовать код с конкретным сжимаемым изображением. Обосновывается выполнимость процедуры сжатия с помощью декодера линейного кода. Показывается, что при кодовом квантовании достигается требуемый уровень сжатия при меньшей сложности (за меньшее время), чем у известных методов векторного квантования.

Ключевые слова *— сжатие изображений, векторное квантование, кодовое векторное квантование, кодовая книга.*

Введение

В основе большинства алгоритмов сжатия изображений с потерями лежит процедура разбиения изображения на правильные фигуры — домены. Над доменом производится некоторое спектральное преобразование, результатом которого является таблица спектральных коэффициентов, вносящих различный по значимости вклад в качество изображения, восстановленного с помощью обратного преобразования. Последнее обстоятельство позволяет различным образом (с различной степенью точности) квантовать спектральные коэффициенты, обеспечивая тем самым сжатие домена без серьезной потери в качестве. При этом спектральные коэффициенты квантуются как скалярные величины, а описанное квантование называется скалярным.

Очевидно, при скалярном квантовании домены изображения обрабатываются независимо друг от друга. Между тем многие домены одного изображения схожи, обладают похожей цветовой гаммой, являются частями одного и того же объекта. Кажется очевидным, что использование этих свойств «похожести» доменов может дать выигрыши, если при сжатии обрабатывать каждый домен как единое целое, а не как множество независимых точек. Чтобы лучше понять предпосылки возможных выигрышей, рассмотрим пример.

Имеется стандартное тестовое изображение размером 512×512 точек, каждая точка является одним из 256 оттенков серого. Разобьем изображение на домены размером 8×8 точек. Домен при этом можно представить вектором длины 64 или, что то же самое, точкой в 64-мерном пространстве. Соответственно, всего в таком векторном пространстве будет $256^{64} = 2^{512}$ точек. Само же изображение будет разбито всего на $4096 = 2^{12}$ доменов, и, следовательно, далеко не все точки пространства будут задействованы. Если бы набор доменов из 2^{12} элементов для этого изображения был известен и кодеру, и декодеру, то можно было бы передать каждый домен максимум 12 битами. Более того, при сжатии с потерей качества все похожие домены внутри изображения могут быть заменены на один, приближающий реальные домены изображения с некоторой ошибкой, что позволит использовать еще меньше бит. Например, если набор используемых образцов будет содержать 256 доменов, то на передачу одного домена изображения потребуется максимум 1 байт, при этом сжатие составило бы не менее 16 раз.

Рассматривая домены как векторы и задавая их отображение в «похожие» векторы из множества квантованных (кодовых) векторов, можно сформулировать процедуру, которая носит название «векторное квантование». Сжатие изображений при векторном квантовании достигается за счет того, что множество квантованных векторов (кодовая книга) выбирается существенно меньшим, чем исходное множество векторов. Векторное квантование изображений исследовалось в работах [1–11].

В настоящей работе рассматриваются возможности построения кодовых книг с помощью аппарата теории кодов, исправляющих ошибки. Такое векторное квантование мы в дальнейшем будем называть кодовым квантованием.

Векторное квантование

Общий метод векторного квантования состоит в том, что изображение разбивается на домены, а домены рассматриваются как векторы или точки в многомерном пространстве, затем при сжатии с потерей качества похожие между собой домены заменяются на один образец. Все образцы помещаются в кодовую книгу.

Векторное квантование обеспечивает сжатие домена в $NL/\log(K)$ раз, где N — длина вектора (число точек в домене); L — число бит на символ одного элемента входного вектора; K — число векторов в кодовой книге.

После применения векторного квантования восстановленное изображение может отличаться от исходного. Уровень искажений при этом будет определяться не только степенью сжатия, но и самим набором векторов в кодовой книге: $\{W\} = \{w_1, w_2, \ldots, w_K\}$. Одна и та же кодовая книга может давать хорошие результаты для одного изображения и плохие для другого. Поэтому в традиционных схемах применения векторного квантования для каждого изображения строится своя кодовая книга. Для этой цели обычно используется обобщенный алгоритм Ллойда [1, 9], позволяющий относительно быстро создавать кодовую книгу. Такой подход к векторному квантованию, однако, связан с большими накладными расходами при хранении и передаче сжатого изображения, поскольку вместе с изображением требуется хранить (передавать) и соответствующую кодовую книгу. Степень сжатия при векторном квантовании можно повысить, если отказаться от хранения кодовой книги, например, если использовать одну и ту же книгу для всех изображений.

Будем говорить, что домен a размера $n \times n$ покрывается в коде W с радиусом R, если найдется $w \in W$, для которого сумма квадратов разностей элементов матрицы $w - a$ не превышает R: $\sum (w_i - a_i)^2 \le R$. Тогда любой код можно рассматривать как покрытие некоего множества доменов радиусом R, а процесс векторного квантования — как отображение множества доменов в кодовые слова $w \in W$.

Для кодовой книги (кода), построенной адаптивно при помощи обобщенного алгоритма Ллойда, достигается минимально возможный радиус покрытия, так как кодовая книга строится исключительно для покрытия доменов сжимаемого изображения. Получаемый с помощью алгоритма Ллойда код, как правило, не имеет компактного описания и, следовательно, требует значительных временных и (или) емкостных затрат на передачу или хранение кодовой книги.

Использование универсальной кодовой книги, представляющей покрытие всего пространства, влечет за собой потери в степени сжатия, поскольку универсальная книга имеет большой радиус покрытия. Однако такую книгу не надо передавать, она известна всем абонентам. Кроме того, существуют коды покрытия, имеющие простое математическое описание, что делает незначительными затраты на хранение книги.

Кодовое векторное квантование

Пусть G — некоторый линейный (n, k)-код над полем $GF(q)$, $q = 256$ для того же случая, когда каждая точка является одним из 256 оттенков серого $GF(q)$. Пусть сторона домена изображения будет равна n_d, при этом $n_d n_d = n$.

Если точки покрываемого пространства являются доменами некоего изображения, то сжатие будет состоять в том, что каждый домен из n точек будет заменен на кодовое слово кода G, наиболее близкое к исходному домену в евклидовой метрике. Так как кодовое слово однозначно определяется своей информационной совокупностью (в нашем случае информационная совокупность имеет длину k), то достаточно хранить k элементов для восстановления всего кодового слова. В результате вместо n точек будет храниться k, а сжатие составит n/k раз. Потери качества будут происходить из-за того, что при замене реальных доменов на кодовые слова идеального соответствия не будет. Главным достоинством при этом является то, что сам код в этом случае вообще не передается. Данный метод для случая использования кодов Рида — Соломона описан в работе [10].

С точки зрения сжатия и последующего восстановления радиус полученного покрытия для изображения влияет на качество восстановленного изображения: чем больше радиус, тем хуже объективное качество. На радиус покрытия будут влиять как количество слов в коде G, так и расположение этих слов в пространстве относительно точек покрываемого множества, и, следовательно, чем лучше выбран код, тем лучше будет качество восстановленного изображения. Если использовать линейный (n, k)-код, то его можно компактно хранить, например, в виде генераторной матрицы кода. Тогда для квантования различных изображений можно использовать разные коды и, следовательно, несколько улучшить качество.

Однако, даже несмотря на то, что при использовании помехоустойчивого кода для квантова-

ния изображений сам код не нужно хранить, общее сжатие оказывается часто меньшим по сравнению со сжатием при адаптивном построении кода. Дело в том, что адаптивное построение позволяет построить код существенно меньшей размерности при одинаковом качестве восстановленных изображений. Поэтому возникает задача согласования кода и изображения

Пусть есть некое пространство C и некий код W, полностью покрывающий пространство C радиусом R. Это означает, что для любой точки пространства C найдется кодовое слово кода W, расстояние от которого до точки пространства не превышает R:

$$\exists w \in W: \forall c \in C, d(c, w) \le R,$$

где $d(c, w) = \sum (w_i - c_i)^2$.

Пусть C' — пространство, точками которого являются реальные домены изображения и пусть есть набор обратимых преобразований M, определенный над множеством точек пространства C' так, что результат преобразований из M всегда принадлежит пространству C:

$$\forall c' \in C', m \in M: c = mc', c \in C.$$

Таким образом, любую точку из пространства C' можно при помощи преобразований из M перевести в точку пространства C, где она будет покрыта кодом W с радиусом R, и, следовательно, использование кодового слова из W в качестве квантователя для точек вида $mc', c' \in C'$ не будет давать ошибку квантования больше, чем R, и преобразования смогут подстроить единый код под конкретное изображение.

При квантовании точек пространства возникает задача поиска ближайшего кодового слова для произвольной точки пространства. В общем случае эта задача может быть решена полным перебором всех кодовых слов. Однако такой перебор требует больших временных затрат и не применим к сжатию в реальном масштабе времени. Однако при использовании в качестве кодовой книги кода, исправляющего ошибки, задача поиска ближайшего вектора кодовой книги упрощается.

Будем полагать, что сжимаемый домен c, представляющий собой вектор над некоторым алфавитом, является суммой ближайшего к нему кодового вектора a и вектора ошибки e: $c = a + e$. Тогда задачу векторного квантования можно решать с помощью процедуры декодирования в шаре радиуса R вектора c в коде W.

Теперь общая схема сжатия изображений при помощи помехоустойчивых кодов можно сформулировать так:

— каждый домен x_i изображения отображается при помощи преобразований в точку c_i пространства C, покрытого кодом W;

— в пространстве C выполняется декодирование вектора c_i в коде W, находится ближайшее кодовое слово w_i, принадлежащее коду W, и это слово w_i считается квантователем для исходного домена x_i;

— сохраняется γ_i — информационная совокупность слова w_i. Сжатие от квантования будет состоять в том, что вместо домена x_i будет храниться информационная совокупность γ_i.

Был описан еще один метод согласования изображения и кода, исправляющего ошибки [11]. Метод основан на разбиении изображения на битовые плоскости с последующим использованием при квантовании этих плоскостей кодов с малой плотностью проверок на четность.

Заключение

В работе описаны некоторые схемы векторного квантования изображений, основанные на использовании методов теории помехоустойчивого кодирования для построения кодовых книг. Экспериментальное сравнение кодовых методов векторного квантования со стандартными алгоритмами показывает, что при некотором проигрыше по степени сжатия изображений кодовые методы имеют значительное преимущество по времени их сжатия.

Литература

1. **Li J., Gray R. M., Olsen R.** Joint Image Compression and Classification with Vector Quantization and Two Dimensional Hidden Markov Model // Proc. of the IEEE Data Compression Conf. (DCC), Snowbird, Utah, 1999. P. 23–32.
2. **Hung A. C., Tsern E. K., Meng T. H.** Error-resilient pyramid vector quantization for image compression // IEEE Transactions on Image Processing. Oct. 1998. Vol. 7. P. 1373–1386.
3. **Lin J.-H., Vitter J. S.** Nearly Optimal Vector Quantization via Linear Programming // Proc. of the IEEE Data Compression Conf. (DCC), Snowbird, Utah, 1992. P. 22–31.
4. **Cosman P. C., Oehler K. L., Riskin E. A., Gray R. M.** Using vector quantization for image processing // Proc. of the IEEE. Sept. 1993. Vol. 81. N 9. P. 1326–1341.
5. **Bayazit U., Pearlman W. A.** Variable-Length Constrained Storage Tree-Structured Vector Quantiza-

tion // IEEE Transactions Image Processing. Mar. 1999. Vol. 8. N 3. P. 321–331.
6. **Bradley J. N., Brislawn C. M.** Wavelet transform-vector quantization compression of supercomputer ocean models // Proc. of the IEEE Data Compression Conf. (DCC), Snowbird, Utah, 1993. P. 224–233.
7. **Raffy P., Antonini M., Barlaud M.** Distortion-Rate Models for Entropy-Coded Lattice Vector Quantization // IEEE Transactions. 2000. Vol. 9. N 12. P. 2006–2017.
8. **Garey M. R., Johnson D. S., Witsenhausen H. S.** The complexity of the generalized Lloyd-Max problem // IEEE Transactions Inform. Theory. 1982. Vol. 28. N 2. P. 255–256.
9. **Gersho A., Gray R. M.** Vector Quantization and Signal Compression. — Kluwer Academic Publishers, 1992. — 732 p.
10. **Белоголовый А. В.** Применение кодов, исправляющих ошибки, для сжатия видеоизображений // Вторая Междунар. молодежная школа-семинар БИКАМП'99: тез. докл. /ГУАП. СПб., 1999. С. 119.
11. **Kabatiansky G., Krouk E., Semenov S.** Error Correcting Coding and Security for Data Networks. Analysis of the Super Channel Concept. — John Wiley & Sons, Ltd., 2005. — 278 p.

УВАЖАЕМЫЕ АВТОРЫ!

При подготовке рукописей статей необходимо руководствоваться следующими рекомендациями.

Статьи должны содержать изложение новых научных результатов. Название статьи должно быть кратким, но информативным. В названии недопустимо использование сокращений, кроме самых общепринятых (РАН, РФ, САПР и т. п.).

Объем статьи (текст, таблицы, иллюстрации и библиография) не должен превышать эквивалента в 20 страниц, напечатанных на бумаге формата А4 на одной стороне через 1,5 интервала Word шрифтом Times New Roman размером 13, поля не менее двух сантиметров.

Обязательными элементами оформления статьи являются: индекс УДК, заглавие, инициалы и фамилия автора (авторов), ученая степень, звание (при отсутствии — должность), полное название организации, аннотация и ключевые слова на русском и английском языках, электронные адреса авторов, которые по требованию ВАК должны быть опубликованы на страницах журнала. При написании аннотации не используйте аббревиатур и не делайте ссылок на источники в списке литературы.

Статьи авторов, не имеющих ученой степени, рекомендуется публиковать в соавторстве с научным руководителем, наличие подписи научного руководителя на рукописи обязательно; в случае самостоятельной публикации обязательно предоставляйте заверенную по месту работы рекомендацию научного руководителя с указанием его фамилии, имени, отчества, места работы, должности, ученого звания, ученой степени — эта информация будет опубликована в ссылке на первой странице.

Формулы набирайте в Word, не используя формульный редактор (Mathtype или Equation), при необходимости можно использовать формульный редактор; для набора одной формулы не используйте два редактора; при наборе формул в формульном редакторе знаки препинания, ограничивающие формулу, набирайте вместе с формулой; для установки размера шрифта никогда не пользуйтесь вкладкой Other..., используйте заводские установки редактора, не подгоняйте размер символов в формулах под размер шрифта в тексте статьи, не растягивайте и не сжимайте мышью формулы, вставленные в текст; в формулах не отделяйте пробелами знаки: + = –.

Для набора формул в Word никогда не используйте Конструктор (на верхней панели: «Работа с формулами» — «Конструктор»), т. к. этот ресурс предназначен только для внутреннего использования в Word и не поддерживается программами, предназначенными для изготовления оригинал-макета журнала.

При наборе символов в тексте помните, что символы, обозначаемые латинскими буквами, набираются светлым курсивом, русскими и греческими — светлым прямым, векторы и матрицы — прямым полужирным шрифтом.

Иллюстрации в текст не заверстываются и предоставляются отдельными исходными файлами, поддающимися редактированию:

— рисунки, графики, диаграммы, блок-схемы изготавливаются в векторных программах: Visio 4, 5, 2002–2003 (*.vsd); Coreldraw (*.cdr); Excel; Word; AdobeIllustrator; AutoCad (*.dxf); Компас; Matlab (*.ps, *.pdf или экспорт в формат *.ai);

— фото и растровые — в формате *.tif, *.png с максимальным разрешением (не менее 300 pixels/inch).

Наличие подрисуночных подписей обязательно (желательно не повторяющих дословно комментарии к рисункам в тексте статьи).

В редакцию предоставляются:

— сведения об авторе (фамилия, имя, отчество, место работы, должность, ученое звание, учебное заведение и год его окончания, ученая степень и год защиты диссертации, область научных интересов, количество научных публикаций, домашний и служебный адреса и телефоны, e-mail), фото авторов: анфас, в темной одежде на белом фоне, должны быть видны плечи и грудь, высокая степень четкости изображения без теней и отблесков на лице, фото можно представить в электронном виде в формате *.tif, *.png с максимальным разрешением — не менее 300 pixels/inch при минимальном размере фото 40 × 55 мм;

— экспертное заключение.

Список литературы составляется по порядку ссылок в тексте и оформляется следующим образом:

— для книг и сборников — фамилия и инициалы авторов, полное название книги (сборника), город, издательство, год, общее количество страниц;

— для журнальных статей — фамилия и инициалы авторов, полное название статьи, название журнала, год издания, номер журнала, номера страниц;

— ссылки на иностранную литературу следует давать на языке оригинала без сокращений;

— при использовании web-материалов указывайте адрес сайта и дату обращения.

Более подробно правила подготовки текста с образцами изложены на нашем сайте в разделе «Оформление статей».

Контакты

Куда: 190000, Санкт-Петербург,
Б. Морская ул., д. 67, ГУАП, РИЦ
Кому: Редакция журнала «Информационно-управляющие системы»
Тел.: (812) 494-70-02
Эл. почта: 80x@mail.ru
Сайт: www.i-us.ru

ПЕРСПЕКТИВНЫЕ МОДЕЛИ И МЕТОДЫ ЗАЩИТЫ КОМПЬЮТЕРНЫХ СЕТЕЙ И ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ КИБЕРПРОСТРАНСТВА: ОБЗОР МЕЖДУНАРОДНЫХ КОНФЕРЕНЦИИ MMM-ACNS-2012 И СЕМИНАРА SA&PS4CS 2012

В Санкт-Петербурге с 17 по 20 октября 2012 года успешно прошли 6-я Международная конференция «Математические модели, методы и архитектуры для защиты компьютерных сетей» (MMM-ACNS-2012) и 2-й Международный семинар «Научный анализ и поддержка политик безопасности в киберпространстве» (SA&PS4CS 2012). Оба мероприятия явились крупными международными форумами специалистов по формальным методам, моделям и архитектурным решениям, принимаемым в области обеспечения безопасности компьютерных сетей и киберпространства, а также по противодействию кибертерроризму. Они продемонстрировали высокий интерес к данной тематике не только исследовательских организаций и ученых всего мира, но и правительственных структур и силовых ведомств практически всех развитых стран, а также международных сообществ. Проведение конференции и семинара позволило осуществить плодотворный обмен мнениями между зарубежными и российскими школами в области защиты информационных ресурсов компьютерных сетей и обеспечения кибербезопасности, способствовало более широкому распространению новых идей и дальнейшему укреплению международного научного сотрудничества в этой области.

Конференция MMM-ACNS-2012 и семинар SA&PS4CS 2012 были организованы Санкт-Петербургским институтом информатики и автоматизации Российской академии наук (СПИИРАН) и Университетом Бингхэмптона — государственным университетом штата Нью-Йорк (США). Сопредседателями конференции и семинара были член-корреспондент РАН Р. М. Юсупов, Р. Герклотц (Управление научных исследований ВВС США) и Ч. Холланд (отделение научных исследований ВМС США в Праге, США), сопредседателями программного комитета конференции — И. В. Котенко (СПИИРАН) и В. Скормин (Университет Бингхэмптона).

На конференции MMM-ACNS-2012 было зарегистрировано 80 участников. Статистические данные об их принадлежности к различным областям деятельности таковы: университетская среда — 36; научные организации — 28; коммерческие организации — 8; государственные учреждения — 6.

Пленарные доклады на конференции MMM-ACNS-2012 были сделаны четырьмя приглашенными докладчиками. **А. Ставро** (Университет Джорджа Мейсона, США) рассмотрел угрозы безопасности, вытекающие из новых возможностей смарт-устройств и онлайновых рынков приложений для мобильных устройств. Предлагаемые в докладе подходы к обеспечению безопасности устройств на платформе Android включают анализ исходного кода и двоичных файлов мобильных приложений с использованием «сети уровня ядра» и шифрования данных, а также управление коммуникационными механизмами для синхронизации контента пользователей с компьютерами и другими телефонами, включая обновления операционной системы или приложений через USB. Был рассмотрен механизм аутентификации USB-соединения, названный USBsec. В докладе **Б. Лившица** (Исследовательская лаборатория «Майкрософт», США) рассматривались имеющиеся у компании «Майкрософт» результаты поиска вредоносных программ в Интернете. Освещались результаты трех исследовательских проектов в этой области: Nozzle, Zozzle и Rozzle. Первые два проекта были направлены на разработку детекторов вредоносных программ (Nozzle — детектор во время выполнения программ, Zozzle — статический детектор). Rozzle является методом, обрабатывающим результаты, полученные первыми двумя детекторами. Исследованию механизмов гарантированного обмена информацией на облачных технологиях был посвящен доклад **Л. Кхана** (Университет Техаса, США), в котором автор рассмотрел ориентированную на облачные вычисления систему доверенного разделения информации. Реализованный в ней механизм управления политиками основан на использовании конфигурационных документов формата RDF. Управление RDF-данными осуществляется с помощью механизма SPARQL-запросов, который широко используется в сообществе Semantic Web и считается более выразительным,

чем XML-ориентированные языки. В докладе **Ф. Мартинелли** (Институт информатики и телематики, Италия) обсуждались качественный и количественный подходы к проблеме реализации политики безопасности. При качественном подходе предлагалось перейти от использования безопасных автоматов к операторам контроллера алгебры процессов (process algebra controller operators). При количественном подходе были рассмотрены такие области, как неточная реализация (inexact enforcement) политики безопасности, вероятностное будущее (probabilistic future), стоимость реализации (cost of enforcement) и отслеживание реализации.

Для организации работы секций международным программным комитетом конференции в ходе ее подготовки были рассмотрены 44 доклада, поданные из 12 стран. Наибольшее количество статей поступило из России, США и Франции. Каждая из статей была тщательно проанализирована тремя-четырьмя рецензентами. В результате было отобрано 22 лучших секционных доклада, представляющих 10 стран: Россию, США, Канаду, Мексику, Италию, Францию, Германию, Норвегию, Испанию и ЮАР. Из этих докладов 14 было выбрано для полных презентаций и 8 — для коротких. Программа конференции предусматривала работу семи секций: «Предотвращение, обнаружение и реагирование на вторжения», «Противодействие вредоносному программному обеспечению», «Прикладная криптография и протоколы безопасности», «Разграничение доступа и защита информации», «Управление событиями и информацией безопасности», «Моделирование защиты информации и безопасность облачных вычислений» и «Политики безопасности».

На панельной дискуссии конференции обсуждались современные проблемы и тенденции в области безопасности компьютерных сетей. В обсуждении этих вопросов приняли участие представители США, Германии, России, Италии, Норвегии, Германии и ЮАР, в том числе и авторы пленарных докладов.

Труды конференции MMM-ACNS-2012 опубликованы в сборнике «Computer Network Security: Sixth International Workshop on Mathematical Methods, Models, and Architectures for Computer Network Security, MMM-ACNS 2012, St.-Petersburg, Russia, October 17–19, 2012, Proceedings / Eds. I. Kotenko, V. Scormin. Berlin: Springer-Verlag, 2012. Vol. 7531.»

Важной особенностью прошедшей конференции MMM-ACNS-2012 явилось сбалансированное сочетание результатов, которые, с одной стороны, посвящены математическому обеспечению информационной безопасности, а с другой — обладают высокой практической значимостью для защиты современных компьютерных сетей. В целом конференция получилась достаточно интересной, ее научный уровень соответствовал мировым стандартам. Более детальную информацию можно найти на сайте http://comsec.spb.ru/mmm-acns12/.

На семинаре SA&PS4CS 2012 было заслушано 16 приглашенных докладов, авторы которых представляли шесть стран: Россию — 8, США — 2, Норвегию — 2, по одному — Италию, Германию и ЮАР. Всего было зарегистрировано 46 участников. Выступления были разделены на три секции. Основными темами выступлений семинара являлись обнаружение, распознавание и определение различных видов деятельности злоумышленников, реагирование на атаки и вторжения в киберпространстве, включая информационные операции национального уровня, идентификация новых перспективных технологий, способов, методов и средств обеспечения взаимодействия в области поддержки политик безопасности в киберпространстве.

Прокомментируем сделанные доклады. В приветственном докладе **В. Майорова** (отдел информационной безопасности, противодействия техническим разведкам и развития систем защиты информации Комитета по информатизации и связи Санкт-Петербурга, Россия) было подчеркнуто, что в условиях глобализации мировой экономики вопросами обеспечения безопасности в киберпространстве обеспокоено все мировое сообщество. В докладе **А. Свистунова** (государственное казенное учреждение Ленинградской области «Оператор «электронного правительства», Россия) были рассмотрены вопросы централизации механизмов управления и мониторинга средствами и системами защиты государственных информационных ресурсов органов исполнительной власти Ленинградской области. В докладе **Р. Юсупова** (СПИИРАН, Россия) были представлены особенности проблемы национальной безопасности в условиях информатизации общества при широкомасштабном внедрении информационных и коммуникационных технологий во все сферы человеческой деятельности, в том числе и в обеспечение национальной безопасности. **Ф. Мартинелли** (Институт информатики и телематики, Италия) рассмотрел проблемы информационной безопасности, касающиеся использования мобильных устройств. Выступление **В. Скормина** (Университет Бингхемптона, США) было посвящено формированию кибербезопасности как научной дисциплины, изучающей законы, не привязанные к какой-либо конкретной технологии или атаке. **А. Грушо** (Институт проблем информатики РАН, Россия) рассмотрел некоторые проблемы компьютерной и сетевой безопасности,

касающиеся выявления скрытых каналов в сетевом трафике, поиска уязвимостей в кодах программного обеспечения, анализа безопасности протоколов, и выявления бот-сетей и центров управления ими. **В. Олещук** (Университет Агдера, Норвегия) проанализировал взаимосвязь кибербезопасности и приватности и имеющиеся в этой области проблемы и конфликты интересов на индивидуальном, организационном и национальном уровнях. В докладе **П. Зегжды** (СПбГТУ, Россия) были систематизированы угрозы на средства виртуализации и системы управления виртуализированными вычислительными системами. **Р. Рике** (Институт безопасных информационных технологий Фраунгофера, Германия) рассмотрел подход к построению SIEM-систем (Security Information and Event Management) следующего поколения, обеспечивающий эффективную поддержку управления и идентификации ситуаций по безопасности. **Э. Хатчисон** (T-Systems International, ЮАР) провел анализ директив и норм Евросоюза, регулирующих приватность данных, и сделал акцент на важных особенностях их применения в SIEM-системах. **А. Смирнов** (Национальный институт исследований глобальной безопасности, Россия) привел результаты исследования влияния современных информационно-коммуникационных технологий на глобальную безопасность и охвата глобальными социальными сетями населения различных стран как принципиально нового геополитического явления. Доклад **А. Земцова** (Group-IB, Россия) был посвящен обзору актуальных проблем законодательной и правоприменительной практики в Российской Федерации в области обеспечения безопасности и борьбы с преступностью в киберпространстве. **Л. Кхан** (Университет Техаса, США) представил структурированный подход к применению различных методов интеллектуального анализа информационных потоков, в частности, применения классификации для анализа эволюционирующих потоков. Доклад **П. Ласкова** (Университет Тюбингена, Германия) был посвящен порталу VirusTotal, на котором собрана коллекция вредоносного программного обеспечения и ведется анализ данных обо всех существующих антивирусных решениях. Доклад **С. Мьёлснеса** (Норвежский университет науки и технологий, Норвегия) был посвящен проблеме семантических атак на протоколы, т. е. атак, которые используют уязвимости в сообщениях и поведении протокола. **И. Котенко** (СПИИРАН, Россия) в своем докладе предложил подход для исследования и реализации различных видов адаптивных и кооперативных способов функционирования команд программных интеллектуальных агентов для реализации распределенных компьютерных атак и механизмов защиты в сети Интернет.

На круглом столе, завершающем работу семинара, было проведено обсуждение форм международного взаимодействия по предупреждению, обнаружению и реагированию на кибервторжения и атаки. Участвовали в работе круглого стола В. Скормин (ведущий), А. Грушо, П. Зегжда, А. Земцов, Л. Кхан, П. Ласков, Ф. Мартинелли, С. Мьёлснес и А. Смирнов.

Прошедший семинар показал стремление всех его участников к объединению усилий в различных областях деятельности, относящихся к научному анализу и поддержке политик безопасности в киберпространстве, а также к активному обмену идеями и изучению последних исследований и разработок в этой важной сфере. В целом семинар был оценен участниками как достаточно интересный и плодотворный. Его научный уровень соответствовал мировым стандартам. Более детальную информацию можно найти на сайте http://www.comsec.spb.ru/saps4cs12/.

Доктор технических наук, профессор,
заведующий лабораторией проблем
компьютерной безопасности СПИИРАН
И. В. Котенко
Доктор технических наук, профессор,
ведущий научный сотрудник
лаборатории проблем компьютерной
безопасности СПИИРАН
И. Б. Саенко

VI МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ «АКУСТООПТИЧЕСКИЕ И РАДИОЛОКАЦИОННЫЕ МЕТОДЫ ИЗМЕРЕНИЙ И ОБРАБОТКИ ИНФОРМАЦИИ» ARMIMP-2013

16–18 сентября 2013 г.

Место проведения: Туристический центр, г. Суздаль
Адрес: ул. Коровники, д. 45, г. Суздаль, Владимирская обл., Россия

На конференции планируется обсудить современные проблемы, связанные с генерированием, излучением, распространением излучения, акустооптическими и радиолокационными методами цифровой обработки сигналов различной физической природы, а также их применение в разных областях науки и техники.

Организаторы
Научно-технологический центр уникального приборостроения РАН
Институт радиотехники и электроники им. В. А. Котельникова РАН
РНТОРЭС им. А. С. Попова
Российская секция IEEE
МГТУ им. Н. Э. Баумана
Владимирский государственный университет
Московский авиационный институт (ГТУ)
ОАО «НПК «Системы прецизионного приборостроения»
ОАО «Концерн радиостроения «Вега»

Направления работы
Генерирование и излучение сверхширокополосных сигналов и сверхкоротких импульсов
Прием, измерение и обработка сверхширокополосных сигналов и сверхкоротких импульсов
Распространение сверхширокополосных сигналов и сверхкоротких импульсов в природных средах
Зондирование природных сред сверхширокополосными сигналами и сверхкороткими импульсами
Миллиметровые и субмиллиметровые волны
Методы математического моделирования физических процессов в оптике и радиолокации
R-функции, атомарные функции, вейвлеты, фракталы и хаос
Информационно-измерительные оптические и радиотехнические системы
Сверхширокополосные хаотические сигналы в оптических и радиотехнических информационных системах
Методы вычислительной томографии в оптике и радиолокации
Лазерная физика и техника
Сверхкороткие импульсы и методы нелинейной оптики
Акустооптические и радиолокационные методы измерений и обработки в биологии, медицине
Физические основы приборостроения

Рабочие языки конференции
Русский и английский

Издание трудов
Принятые доклады будут опубликованы в трудах конференции.
Сборник трудов конференции выйдет из печати к началу конференции. Докладчики и слушатели, оплатившие целевой взнос, получат сборник трудов и программу конференции при регистрации.
По решению программного комитета отдельные работы будут опубликованы в специальных англоязычных выпусках международных журналов «Электромагнитные волны и электронные системы», «Физические основы приборостроения».

Контрольные сроки
Прием докладов до 10 июля 2013 г.
Все материалы предоставляются в электронном виде по адресу: doklad-rntores@mailru
Формат файлов: только MS Word doc (любой версии). В формате PDF тезисы не принимаются.

Дополнительная информация и справки
Адрес программного комитета:
107031, г. Москва, Российское НТОРЭС им. А. С. Попова, Рождественка, 6/9/20, стр. 1
doklad-rntores@mailru
http://wwwrntoresru
Тел: +7(495) 621-71-08
Факс: +7(495) 621-06-10
Ученый секретарь РНТОРЭС Геннадий Третьяков

СВЕДЕНИЯ ОБ АВТОРАХ

АЛФЁРОВА Наталья Васильевна

Доцент кафедры теории общественного развития стран Азии и Африки Санкт-Петербургского государственного университета.
В 1997 году окончила Санкт-Петербургский государственный университет по специальности «Историк искусств».
В 2006 году защитила диссертацию на соискание ученой степени кандидата культурологии.
Является автором десяти научных публикаций.
Область научных интересов — история и теория искусств, история искусств стран Ближнего Востока, культурология.
Эл. адрес: alferovan@yandex.ru

ВИНЕЛЬ Алексей Викторович

Ведущий научный сотрудник НПФ «Информационные и сетевые технологии», г. Москва.
В 2005 году окончил Санкт-Петербургский государственный университет аэрокосмического приборостроения по магистерскому направлению «Информационные системы в экономике».
В 2007 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором более 50 научных публикаций.
Область научных интересов — случайный множественный доступ, анализ и оценка производительности беспроводных сетей передачи данных.
Эл. адрес: vinel@ieee.org

ГРИГОРЬЕВЫХ Елена Андреевна

Ассистент кафедры радиотехнических и медико-биологических систем Поволжского государственного технического университета, г. Йошкар-Ола.
В 2007 году окончила Марийский государственный технический университет по направлению «Радиосвязь, радиовещание и телевидение».
Является автором 17 научных публикаций.
Область научных интересов — системы передачи информации, цифровая обработка сигналов.
Эл. адрес: GrigorevyhEA@volgatech.net

АНДРЕЕВ Николай Дмитриевич

Начальник отдела разработки ПО компании «Джи Джи Эй», ассистент кафедры прикладной математики Санкт-Петербургского государственного политехнического университета.
В 2002 году окончил физико-механический факультет Санкт-Петербургского государственного политехнического университета по специальности «Прикладная математика и информатика».
Является автором трех научных публикаций.
Область научных интересов — процессы разработки программного обеспечения, предметно-ориентированные языки, технологии программирования.
Эл. адрес: nikolai.andreev@gmail.com

ГОРОДЕЦКИЙ Андрей Емельянович

Доктор технических наук, профессор, заведующий лабораторией методов и средств автоматизации Института проблем машиноведения РАН, г. Санкт-Петербург, заслуженный деятель науки и техники.
В 1965 году окончил Ленинградский политехнический институт им. М. И. Калинина.
В 1993 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 150 научных публикаций и 70 изобретений.
Область научных интересов — математическое моделирование, оптимальное управление, идентификация и диагностика.
Эл. адрес: gorodetsky@mail23.ipme.ru

ДЕМЬЯНЧУК Анна Алексеевна

Младший научный сотрудник научно-исследовательского отдела проблем информационной безопасности Санкт-Петербургского института информатики и автоматизации РАН, аспирант Санкт-Петербургского государственного электротехнического университета «ЛЭТИ».
В 2011 году окончила Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» по специальности «Компьютерная безопасность».
Является автором двух научных публикаций.
Область научных интересов — информационная безопасность, математические основы криптографии, двухключевая криптография, электронные цифровые подписи, протоколы с нулевым разглашением.
Эл. адрес: anonimkina@gmail.com

ДО
Чо
Суан

Гражданин Вьетнама.
Аспирант кафедры САПР Санкт-Петербургского государственного электротехнического университета «ЛЭТИ».
В 2010 году окончил магистратуру Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» по специальности «Информатика и вычислительная техника».
Область научных интересов — моделирование, системы управления, алгоритмизация.
Эл. адрес:
doxuancholeti@yahoo.com

ЗЕЛЕНЕВ
Евгений
Ильич

Профессор кафедры теории общественного развития стран Азии и Африки Санкт-Петербургского государственного университета, почетный академик Академии наук республики Башкортостан, почетный профессор Ханойского национального университета (Вьетнам), Ереванского государственного университета (Армения).
В 1979 году окончил Ленинградский государственный университет по специальности «Востоковед-историк».
В 2000 году защитил диссертацию на соискание ученой степени доктора исторических наук.
Является автором 100 научных публикаций и десяти монографий.
Область научных интересов — история стран Ближнего и Среднего Востока, исламоведение, культурология.
Эл. адрес: evzelenev@gmail.com

КОТЕНКО
Игорь
Витальевич

Профессор, заведующий лабораторией проблем компьютерной безопасности Санкт-Петербургского института информатики и автоматизации РАН.
В 1983 году окончил Военно-космическую академию им. А. Ф. Можайского по специальности «Математическое обеспечение автоматизированных систем управления», в 1987 году — Военную академию связи по специальности «Инженерная автоматизированных систем управления».
В 1999 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 450 научных публикаций.
Область научных интересов — безопасность компьютерных сетей, обнаружение компьютерных атак, межсетевые экраны, защита от вирусов и сетевых червей и др.
Эл. адрес: ivkote@comsec.spb.ru

ЗАБРОВСКИЙ
Анатолий
Леонидович

Аспирант математического факультета, ведущий программист Петрозаводского государственного университета.
В 2008 году окончил магистратуру Петрозаводского государственного университета по специальности «Информатика и вычислительная техника».
Является автором восьми научных публикаций.
Область научных интересов — мультимедийные сетевые технологии, качество передачи мультимедийных потоков.
Эл. адрес: z_anatoliy@petrsu.ru

ЗИНЯКОВ
Владимир
Юрьевич

Аспирант кафедры систем и технологий управления Института информационных технологий и управления Санкт-Петербургского государственного политехнического университета.
В 2012 году окончил Санкт-Петербургский государственный политехнический университет по специальности «Информатика и вычислительная техника».
Область научных интересов — системы автоматизированного управления, обработка изображений.
Эл. адрес: vziniakov@gmail.com

КРАСИЛЬНИКОВ
Николай
Николаевич

Профессор кафедры информационно-сетевых технологий Санкт-Петербургского государственного университета аэрокосмического приборостроения, заслуженный деятель науки и техники РФ.
В 1950 году окончил Ленинградский политехнический институт по специальности «Техническая физика».
В 1963 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 240 научных публикаций.
Область научных интересов — цифровая обработка изображений, статистическая теория передачи и восприятия изображений, математическое моделирование процессов обработки информации зрительной системой человека.
Эл. адрес:
NNKrasilnikov@yandex.ru

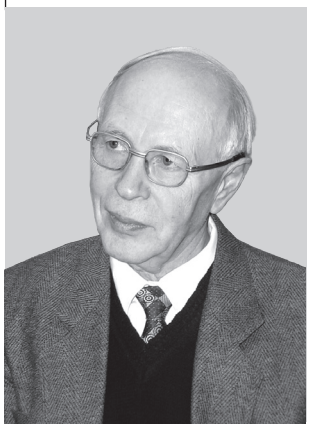
КРУК
Евгений
Аврамович

Профессор, заведующий кафедрой безопасности информационных систем, декан факультета информационных систем и защиты информации Санкт-Петербургского государственного университета аэрокосмического приборостроения, заслуженный деятель науки РФ, лауреат премии Правительства Санкт-Петербурга.
В 1974 году окончил Ленинградский институт аэрокосмического приборостроения по специальности «Автоматические системы управления».
В 1999 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 120 научных публикаций.
Область научных интересов — теория кодирования, криптография.
Эл. адрес: ekrouk@vu.spb.ru

КУЧМИН
Андрей
Юрьевич

Старший научный сотрудник лаборатории механики управляемых систем Института проблем машиноведения РАН, г. Санкт-Петербург.
В 2005 году окончил Санкт-Петербургский государственный университет аэрокосмического приборостроения.
В 2007 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором 22 научных публикаций и двух патентов на изобретения.
Область научных интересов — математическое моделирование в естественных науках, искусственный интеллект и принятие решений, математические проблемы теории управления и др.
Эл. адрес:
radiotelescope@yandex.ru

МЕЛЕХИН
Виктор
Федорович



Профессор, заведующий кафедрой компьютерных систем и программных технологий Санкт-Петербургского государственного политехнического университета, почетный работник высшего профессионального образования РФ.
В 1960 году окончил Ленинградский политехнический институт.
В 1984 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 230 научных публикаций, в том числе четырех монографий и 75 изобретений.
Область научных интересов — теория и технология проектирования вычислительных систем и устройств.
Эл. адрес:
melekhin@kspt.ftk.spbstu.ru

КУРИЛОВА
Оксана
Леонидовна

Старший преподаватель кафедры телекоммуникационных технологий и сетей, аспирант Ульяновского государственного университета.
В 1994 году окончила Московский государственный университет им. М. В. Ломоносова (филиал в г. Ульяновске) по специальности «Прикладная математика».
Является автором восьми научных публикаций и шести учебно-методических пособий.
Область научных интересов — методы оптимизации, генетические алгоритмы, web-программирование, теория графов, мировые информационные ресурсы и сети.
Эл. адрес:
oxana197208@rambler.ru

МАКСИМЕНКО
Сергей
Леонидович

Старший преподаватель кафедры компьютерных систем и программных технологий Санкт-Петербургского государственного политехнического университета.
В 1998 году окончил с отличием Санкт-Петербургский государственный технический университет по специальности «Вычислительные машины, комплексы, системы и сети».
Является автором более 20 научных публикаций.
Область научных интересов — технологии проектирования аппаратуры вычислительных систем, системы на кристалле, цифровая обработка сигналов.
Эл. адрес:
sl_max@kspt.ftk.spbstu.ru

МИРИН
Анатолий
Юрьевич

Старший научный сотрудник научно-исследовательского отдела проблем информационной безопасности Санкт-Петербургского института информатики и автоматизации РАН.
В 2002 году окончил Санкт-Петербургский государственный университет аэрокосмического приборостроения по специальности «Вычислительные машины, комплексы, системы и сети».
В 2005 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором 18 научных публикаций.
Область научных интересов — информационная безопасность, криптографические протоколы, блочные шифры, хэш-функции.
Эл. адрес: mirin@cobra.ru

МОЛДОВЯН
Николай
Андреевич

Профессор, заведующий научно-исследовательским отделом проблем информационной безопасности Санкт-Петербургского института информатики и автоматизации РАН, заслуженный изобретатель РФ.
В 1975 году окончил Кишиневский политехнический институт по специальности «Полупроводниковые приборы».
В 2001 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 250 научных публикаций и 60 патентов на изобретения.
Область научных интересов — информационная безопасность, криптография, электронная цифровая подпись, блочные шифры.
Эл. адрес: nmold@mail.ru

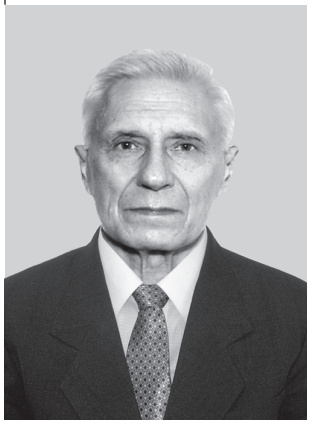
НОВИКОВ
Евгений
Александрович

Докторант кафедры сетей и систем связи космических комплексов Военно-космической академии им. А. Ф. Можайского, г. Санкт-Петербург.
В 2002 году окончил Военный инженерно-космический университет им. А. Ф. Можайского по специальности «Автоматизированные системы обработки информации и управления».
В 2007 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором более 30 научных публикаций.
Область научных интересов — оптимальное управление динамическими системами, управление ресурсами телекоммуникационной системы.
Эл. адрес: moozg@mail.ru

НОВИКОВ
Федор
Александрович

Профессор кафедры прикладной математики Санкт-Петербургского государственного политехнического университета и кафедры компьютерных технологий Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики.
В 1974 году окончил Ленинградский государственный университет по специальности «Математик».
В 2011 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором 74 научных публикаций.
Область научных интересов — прикладная математика, визуальное моделирование программного обеспечения, технологии программирования.
Эл. адрес:
fedornovikov51@gmail.com

МОСКАЛЕЦ
Олег
Дмитриевич



Доцент кафедры электроники и оптической связи Санкт-Петербургского государственного университета аэрокосмического приборостроения.
В 1961 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Радиотехника».
В 1970 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором 120 научных публикаций и пяти патентов на изобретения.
Область научных интересов — теория сигналов, теория линейных систем, спектрально-корреляционный анализ сигналов, квантовая физика.
Эл. адрес: molegd@mail.ru

НОВИКОВА
Евгения
Сергеевна

Ассистент кафедры автоматизированных систем управления и обработки информации Санкт-Петербургского государственного электротехнического университета «ЛЭТИ».
В 2007 году окончила с отличием Санкт-Петербургский электротехнический университет «ЛЭТИ» по специальности «Компьютерная безопасность».
В 2009 году защитила диссертацию на соискание ученой степени кандидата технических наук.
Является автором 46 научных публикаций и трех патентов на изобретения.
Область научных интересов — информационная безопасность, визуализация событий безопасности, двухключевая криптография, электронные цифровые подписи и др.
Эл. адрес:
novikova@comsec.spb.ru

РОГОВ
Александр
Александрович

Профессор, заведующий кафедрой теории вероятностей и анализа данных Петрозаводского государственного университета, почетный работник высшего профессионального образования РФ.
В 1985 году окончил Петрозаводский государственный университет по специальности «Математика».
В 2004 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором 123 научных публикаций.
Область научных интересов — математическое моделирование, математическая и прикладная статистика, интеллектуальные информационные системы.
Эл. адрес: rogov@psu.karelia.ru

САЕНКО Игорь Борисович

Профессор, ведущий научный сотрудник лаборатории проблем компьютерной безопасности Санкт-Петербургского института информатики и автоматизации РАН.
В 1981 году окончил Белорусский государственный университет по специальности «Радиофизика и электроника».
В 2001 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором 150 научных публикаций и семи патентов на изобретения.
Область научных интересов — автоматизированные информационные системы, компьютерная безопасность, методы эволюционного моделирования, теория баз данных.
Эл. адрес: ibsaen@comsec.spb.ru

ТАРАНИН Владимир Валерьевич

Аспирант кафедры информатики и информационной безопасности Петербургского государственного университета путей сообщения.
В 2012 году окончил Петербургский государственный университет путей сообщения по специальности «Комплексное обеспечение информационной безопасности автоматизированных систем».
Область научных интересов — системы защиты информации, криптографические протоколы, защита информации в сетях передачи данных.
Эл. адрес: vovataranin@mail.ru

ЧЕРНОВ Владимир Георгиевич

Профессор кафедры управления и информатики в технических и экономических системах Владимирского государственного университета.
В 1966 году окончил Рязанский радиотехнический институт.
В 1971 году защитил диссертацию на соискание ученой степени кандидата технических наук, в 2007 году — доктора экономических наук.
Является автором более 115 научных публикаций, трех монографий и 15 патентов на изобретения.
Область научных интересов — системы и методы поддержки принятия решений для слабоструктурированных задач, приложения аппарата нечетких множеств в исследованиях экономических процессов.
Эл. адрес: Vladimir.chernov@rambler.ru

СОЛЬНИЦЕВ Ремир Иосифович

Профессор кафедры компьютерного проектирования информационно-измерительных систем Санкт-Петербургского государственного университета аэрокосмического приборостроения.
В 1956 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Гироскопические приборы и системы стабилизации».
В 1970 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором 300 научных публикаций.
Область научных интересов — системы автоматизации проектирования, системы управления, экологические системы.
Эл. адрес: remira70@mail.ru

ХАФИЗОВ Ринат Гафиятуллович

Профессор кафедры радиотехнических и медико-биологических систем Поволжского государственного технологического университета, г. Йошкар-Ола.
В 1994 году окончил Марийский государственный технический университет по специальности «Радиотехника».
В 2010 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 100 научных публикаций.
Область научных интересов — обработка изображений групповых точечных объектов, цифровая обработка сигналов, системы передачи информации.
Эл. адрес: HafizovRG@volgatech.net

ЮРКИН Дмитрий Валерьевич

Доцент кафедры информационной безопасности телекоммуникационных систем Санкт-Петербургского государственного университета телекоммуникаций им. проф. М. А. Бонч-Бруевича.
В 2006 году окончил Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича по специальности «Защищенные системы связи».
Является автором более 30 учебных и научных публикаций.
Область научных интересов — системы радиосвязи и защиты информации, криптографические протоколы, методы согласования работы дискретных автоматов.
Эл. адрес: DVYurkin@yandex.ru

УДК 004.932

Метод формирования 3D-изображения сцены по одной фотографии

Красильников Н. Н. Информационно-управляющие системы, 2013. № 3. С. 2–7.

Описан метод формирования 3D-изображения сцены, основанный на аппроксимации ее центральной проекции (2D-изображения) набором плоскостей с последующей декомпозицией сцены на эти плоскости. Для каждой из плоскостей находится аксонометрическая проекция путем использования имеющейся априорной информации об изображенной сцене и карта глубины. Завершающим шагом описываемого метода является «сборка» 3D-изображения сцены объединением аппроксимирующих ее плоскостей и проекция сцены на экран при заданных условиях наблюдения.

Ключевые слова — 3D-изображение, 3D-сканирование, карта глубины.

Список лит.: 10 назв.

UDC 004.932

The Method of Generating a 3D-Image Scene Based on a Single Photograph

Krasilnikov N. N. IUS, 2013. N 3. P. 2–7.

There has been described the method of generating a 3D-image of a scene based on approximation of the scene central projection (2D-image) by a set of planes with subsequent decomposition of a scene into these planes. An isometric view is defined for each plane by using a priori information of the scene and the depth map. The final stage of this method is assembling of a 3D-image scene by merging the planes which approximate it and projecting the scene on the screen under given conditions of observation.

Keywords — 3D-Image, 3D-Scanning, Depth Map.

Refs: 10 titles.

УДК 621.391

Восстановление двумерных изображений с дефектами

Зиняков В. Ю., Городецкий А. Е., Кучмин А. Ю., Зеленев Е. И., Алферова Н. В. Информационно-управляющие системы, 2013. № 3. С. 8–15.

Рассматривается задача обработки и восстановления изображений типа графический орнамент. Предложен новый подход, который заключается в комбинировании известных математических алгоритмов для достижения бóльших надежности и вычислительной экономичности алгоритма восстановления. Метод проходил тестирование и был оптимизирован для анализа и последующего восстановления исторических орнаментов, составленных главным образом из геометрических мотивов.

Ключевые слова — обработка изображений, вейвлет-анализ, фрактальный анализ.

Список лит.: 5 назв.

UDC 621.391

Restoration of 2D Defective Images

Ziniakov V. Yu., Gorodetskiy A. E., Kuchmin A. Yu., Zelenev E. I., Alferova N. V. IUS, 2013. N 3. P. 8–15.

There has been considered a task of images processing and restoration. An innovative approach is presented which suggests combining known mathematical algorithms to achieve higher reliability and computational efficiency of a restoration algorithm. The method was tested and optimized for analysis and subsequent restoration of historical ornaments mainly composed of geometrical patterns.

Keywords — Image Processing, Wavelet Analysis, Fractal Analysis.

Refs: 5 titles.

УДК 681.3

Анализ надежности цифровых устройств со структурным резервированием и периодическим восстановлением работоспособного состояния узлов

Максименко С. Л., Мелехин В. Ф. Информационно-управляющие системы, 2013. № 3. С. 16–22.

Предлагается метод оценки надежности цифровых устройств, представленных в виде сети троированных узлов, учитывающий цикличность вычислительных процессов и периодическое восстановление информации в узлах. Метод основан на разбиении устройства на ячейки с независимыми отказами. Для произвольной ячейки предложена приближенная оценка вероятности безотказной работы для случая, когда период восстановления информации в узлах много меньше среднего интервала между отказами. Получена зависимость интенсивности отказов ячейки от периода восстановления информации и показано, что эта зависимость линейная независимо от структуры связей в ячейке.

Ключевые слова — цифровые устройства, интегральные схемы, сбои, отказы, периодическое восстановление, надежность, резервирование, оценка, модель, структура, троирование, мажоритар.

Список лит.: 5 назв.

UDC 681.3

Reliability Analysis of Digital Devices with Structural Redundancy and Periodic Operational State Recovery of Functional Nodes.

Maximenko S. L., Melekhin V. F. IUS, 2013. N 3. P. 16–22.

There has been proposed a method to estimate reliability of digital devices represented as nets of triplicated nodes, which takes into consideration cyclic properties of computational process and periodic information recovery in the nodes. The method is based on partitioning a device structure into cells with independent failures. For a random cell there has been given approximate estimation of reliability probability for a case when information recovery period in the nodes is much less than average failure cycle. There has been obtained a relation of cell failure rate to information recovery period and shown that this relation is linear regardless communication structure in a cell.

Keywords — Digital Devices, Integrated Circuit, Fault, Soft Error, Periodic Recovery, Reliability, Structural Redundancy, Model, Structure, Triplication, Voter.

Refs: 5 titles.

УДК 658.562.3

Модификация алгоритмов управления, использующих правила нечеткого условного вывода

Чернов В. Г. Информационно-управляющие системы, 2013. № 3. С. 23–29.

Анализируются недостатки известных алгоритмов управления на основе правил нечеткого условного вывода. Предлагается модификация алгоритмов, устраняющая выявленные недостатки. Представлены результаты моделирования, показывающие, что новый подход позволяет получить лучшее качество управления.

Ключевые слова — нечеткое множество, функция принадлежности, правила нечеткого условного вывода.

Список лит.: 11 назв.

UDC 658.562.3

Modification of Control Algorithms Using Rules of Fuzzy Conditional Conclusion

Chernov V. G. IUS, 2013. N 3. P. 23–29.

There have been analyzed shortcomings of well-known control algorithms using rules of fuzzy conditional conclusion. There has been presented algorithms modification which eliminates identified shortcomings. The simulation results presented show that the new approach allows a better quality of control.

Keywords — Fuzzy Set, Membership Function, the Rules of Fuzzy Conditional Conclusion.

Refs: 11 titles.

УДК 658.512.22

Алгоритмизация обработки и передачи метеорологических данных в замкнутой системе управления «Природа-техногеника»

Сольницев Р. И., До Суан Чо. Информационно-управляющие системы, 2013. № 3. С. 30–35.

Излагается дальнейшее развитие теории и практики создания замкнутой системы управления «Природа-техногеника», предназначенной для эффективного снижения загрязняющих веществ, выбрасываемых промышленными предприятиями в атмосферу. Представлены алгоритмы обработки и передачи метеорологических данных в систему управления «Природа-техногеника».

Ключевые слова — экология, загрязняющие вещества, система автоматического управления, метеорологическое обеспечение, алгоритм.

Список лит.: 8 назв.

UDC 658.512.22

Algorithmization of Meteorological Data Processing and Transmission within a Closed Control System «Nature-Technogenics»

Solnitsev R. I., Do Xuan Cho. IUS, 2013. N 3. P. 30–35.

There has been described further development of theory and practice creating the closed control system «Nature-Technogenics» aimed at effective reducing of pollutants emitted into the atmosphere by industrial enterprises. Algorithms for processing the meteorological information and transmitting it to the control system «Nature-Technogenics» are presented.

Keywords — ecology, pollutants, automatic control system, modeling, meteorological support, algorithm.

Refs: 8 titles.

УДК 621.396.96

Модели сигналов в радиополяриметрии

Москалец О. Д. Информационно-управляющие системы, 2013. № 3. С. 36–41.

Предлагается векторная модель сигналов при исследовании поляризационных характеристик электромагнитных волн при их классическом описании. Введены поляризационные спектры векторных сигналов, где каждая бесконечно малая векторная монохроматическая компонента имеет свое, индивидуальное состояние поляризации. Установлено существование границ классического, приближенного, описания электромагнитных сигналов во временной и частотной областях. Показана необходимость развития физического аспекта теории сигналов.

Ключевые слова — информация, сигнал, скалярная модель сигнала, векторная модель сигнала, состояние поляризации, поляризационный спектр, вектор Джонса, физический аспект.

Список лит.: 21 назв.

UDC 621.396.96

Signal Models in Radiopolarimetry

Moskaletz O. D. IUS, 2013. N 3. P. 36–41.

There has been provided a signal vector model during electromagnetic waves polarization characteristics research using their classical description. There have been introduced polarization spectra of vector signals where every infinitely small vector monochromatic component has its individual polarization state. There has been established existence of bounds of classical approximate description of electromagnetic signals in time and frequency domains. There has been demonstrated a necessity to develop a physical aspect of the signal theory.

Keywords — Information, Signal, Scalar Signal Model, Vector Signal Model, Polarization State, Polarization Spectrum, Jones Vector, Physical Aspect.

Refs: 21 titles.

УДК 004.9

Система моделирования сетевых помех мультимедийных потоков

Рогов А. А., Забровский А. Л. Информационно-управляющие системы, 2013. № 3. С. 42–46.

Представлена система моделирования сетевых помех, влияющих на качество мультимедийных потоков, передаваемых в IP-сети. Данная система предназначена для исследования и тестирования новых мультимедийных сервисов и систем, а также проверки создаваемых критериев оценки качества мультимедийных потоков. Показано, что для определения результирующего качества мультимедийных потоков, передаваемых в реальном режиме времени, можно использовать параметры, полученные от плееров удаленных пользователей. Для оценки качества использовались время начала воспроизведения, минимальное количество кадров в секунду, максимальный скачок потери кадров и минимальный размер буфера в секундах, который был зафиксирован в течение всего воспроизведения мультимедийного потока.

Ключевые слова — система, моделирование, сетевые помехи, эмулятор, WANem, сеть, качество, мультимедийный поток.

Список лит.: 6 назв.

UDC 004.09

Simulation System of Network Disturbances of Multimedia Streams

Rogov A. A., Zabrovskiy A. L. IUS, 2013. N 3. P. 42–46.

There has been presented a simulation system of network disturbances affecting quality of multimedia streams transmitted in an IP network. The system has been designed for investigating and testing new multimedia services and systems and checking developed criteria of quality assessment of multimedia streams. It is shown that to determine resulting quality of multimedia streams transmitted in the real time mode parameters received from players of remote users can be used. Such parameters as playback start time, minimum number of frames per second, maximum jump (number) of frame loss, and minimum buffer size in seconds which has been registered during an entire playback of a multimedia stream have been used for quality assessment.

Keywords — System, Simulation, Network Disturbances, Emulator, Wanem, Network, Quality, Multimedia Stream.

Refs: 6 titles.

УДК 004.434

Фабрики прикладного программного обеспечения, управляемые моделями предметных областей

Андреев Н. Д., Новиков Ф. А. Информационно-управляющие системы, 2013. № 3. С. 47–54.

Обсуждаются методы повышения продуктивности разработки прикладного программного обеспечения на основе определения и использования моделей предметных областей и языков предметной области. Предлагаются принципы применения разработки, управляемой моделью предметной области, для создания фабрик прикладного программного обеспечения. Приводится пример разработанного языка предметной области и указываются преимущества, которые дает его использование.

Ключевые слова — разработка программного обеспечения, фабрики программного обеспечения, управляемая моделью разработка, предметно-ориентированные языки.

Список лит.: 23 назв.

UDC 004.434

Domain Model Driven Applied Software Factories

Andreev N. D., Novikov F. A. IUS, 2013. N 3. P. 47–54.

There have been discussed methods of improving productivity of software development based on definition and use of domain models and domain specific languages. Principles of applying domain model driven development to build applied software factories have been proposed. There has been given an example of a developed domain specific language, its benefits have been described.

Keywords — Software Development, Software Factories, Model Driven Development, Domain Specific Languages.

Refs: 23 titles.

УДК 681.3.06 (075.8)

Визуальный анализ защищенности компьютерных сетей

Котенко И. В., Новикова Е. С. Информационно-управляющие системы, 2013. № 3. С. 55–61.

Исследуются методики визуального анализа защищенности компьютерных сетей. Описывается компонент визуализации системы оценки защищенности компьютерной сети, отличающийся от других систем тем, что позволяет графически представлять как отчеты сканеров уязвимостей, так и результаты моделирования атак, благодаря чему пользователь системы может соотнести потенциальные причины нарушения безопасности с возможными последствиями их эксплуатации злоумышленником.

Ключевые слова — визуализация событий безопасности, оценка защищенности, политики безопасности, графы атак, карты деревьев.

Список лит.: 21 назв.

UDC 681.3.06 (075.8)

Visual Analysis of Computer Network Security Assessment

Kotenko I. V., Novikova E. S. IUS, 2013. N 3. P. 55–61.

There have been studied different visualization techniques for assessment of computer network security. Visualization component of a network security evaluation system has been described; it differs from other systems by its ability to visualize both reports of scanner vulnerability and attack graphs, thus, providing the system operator with an opportunity to correlate potential causes of breach of security and possible consequences of the system utilization by an intruder.

Keywords — Security Visualization, Security Evaluation, Security Policies, Attack Graphs, Treemaps.

Refs: 21 titles.

УДК 004.05

Анализ временных и сложностных характеристик парольной аутентификации в защищенных операционных системах семейства Unix

Юркин Д. В., Винель А. В., Таранин В. В. Информационно-управляющие системы, 2013. № 3. С. 62–66.

Описан подход к оценке вероятностно-временных характеристик протоколов аутентификации в операционных системах семейства Unix, основывающийся на теории вероятностных графов. Показано влияние действий нарушителя на работу протоколов аутентификации.

Ключевые слова — криптографические протоколы, Unix OS, вероятностные графы.

Список лит.: 5 назв.

UDC 004.05

Analysis of Time and Complexity Characteristics of Password Authentication in Protected Unix Operating Systems

Yurkin D. V., Vinel A. V., Taranin V. V. IUS, 2013. N 3. P. 62–66.

An approach to estimate time-probabilistic characteristics of authentication protocols in Unix OS by means of the random graphs theory has been proposed. Influence of intruder's actions on performance of authentication protocols has been shown.

Keywords — Cryptographic Protocols, Performance Analysis, Unix OS, Probabilistic Graphs.

Refs: 5 titles.

УДК 681.3

Типы и приложения протоколов с нулевым разглашением секрета

Демьянчук А. А., Мирин А. Ю., Молдовян Н. А. Информационно-управляющие системы, 2013. № 3. С. 67–73.

Представлены новые варианты протоколов с нулевым разглашением на основе трудности задач дискретного логарифмирования и факторизации. Обсуждается способ доказательства стойкости схем электронной цифровой подписи построением их путем преобразования протоколов с нулевым разглашением секрета. Предложен ряд новых протоколов с нулевым разглашением, включая двухпроходные.

Ключевые слова — криптографический протокол, аутентификация, открытый ключ, электронная цифровая подпись, задача дискретного логарифмирования, задача факторизации.

Список лит.: 14 назв.

UDC 681.3

Types and Applications of Zero-Knowledge Protocols

Demiyanchuk A. A., Mirin A. Y., Moldovyan N. A. IUS, 2013. N 3. P. 67–73.

There have been presented new variants of zero-knowledge protocols based on complexity of discrete logarithm and factoring problems. There has been discussed a method of proving security of digital signature schemes through their construction as a transformation of zero knowledge protocols. There have been proposed a number of new zero knowledge protocols including two-pass protocols.

Keywords — Cryptographic Protocol, Authentication, Public Key, Digital Signature, Discrete Logarithm Problem, Factoring Problem.

Refs: 14 titles.

УДК 621.391.266

Формирование и обработка комплекснозначных последовательностей в многоканальных системах передачи информации

Григорьевых Е. А., Хафизов Р. Г. Информационно-управляющие системы, 2013. № 3. С. 74–77.

Предложен подход к организации многоканальной передачи информации с использованием комплекснозначных последовательностей. Показано, что применение комплекснозначных последовательностей, обладающих равномерным энергетическим спектром, позволяет устранить влияние соседних каналов. Рассмотрен принцип формирования кадра при многоканальной передаче информации на примере двоичных каналов.

Ключевые слова — многоканальная связь, комплекснозначный сигнал, кодовое разделение каналов, пропускная способность.

Список лит.: 5 назв.

UDC 621.391.266

Generation and Processing of Complex-Valued Sequences in Multichannel Data Transmission Systems

Grigorevykh E. A., Khafizov R. G. IUS, 2013. N 3. P. 74–77.

There has been proposed an approach to organization of multichannel data transmission using complex-valued sequences. It has been shown that use of complex-valued sequences having a uniform energy spectrum allows eliminating influence of neighboring channels. The principle of framing during multichannel data transmission on the example of binary channels has been considered.

Keywords — Multichannel Communication, Complex-Valued Signal, Code Division Access, Bandwidth.

Refs: 5 titles.

УДК 621.396

Применение моделей структурной динамики при решении задачи распределения частотно-временного ресурса сети спутниковой связи на основе стандарта DVB-RCS

Новиков Е. А. Информационно-управляющие системы, 2013. № 3. С. 78–83.

Рассмотрен стандарт использования спутникового ресурса DVB-RCS, в частности структура частотно-временного ресурса «обратных» каналов спутников-ретрансляторов. Определены основные недостатки используемых алгоритмов решения задачи оперативного распределения ресурса спутника-ретранслятора. Сформулирована и решена задача оптимального планирования ресурса «обратного» канала на основе моделей структурной динамики.

Ключевые слова — спутниковая связь, DVB-RCS, MF-TDMA, обратный канал, частотно-временной ресурс.

Список лит.: 24 назв.

UDC 621.396

Application of Structural Dynamics Models in Tasks of Time-Frequency Source Distribution in a Satellite Communication System based on DVB-RCS Standard

Novikov E. A. IUS, 2013. N 3. P. 78–83.

There has been considered a standard of application of a satellite source DVB-RCS, in particular, a structure of time-frequency source of return channels in satellites-retransmitters. There have been defined general disadvantages of used algorithms for operative distribution of sources of a satellite-retransmitter. There has been formulated and solved a task of an optimal return channel source based structural dynamics models.

Keywords — Satellite Communication, DVB-RCS, MF-TDMA, Return Channel, Time-And-Frequency Resource.

Refs: 24 titles.

УДК 004.02:378

Применение генетического алгоритма для оптимизации учебного плана

Курилова О. Л. Информационно-управляющие системы, 2013. № 3. С. 84–92.

Представлен алгоритм оптимизации учебного плана в рамках компетентностного подхода. Разработан алгоритм построения матрицы смежности дисциплин, алгоритм ориентированного графа дисциплин, алгоритм нахождения самого длинного пути в графе. Продемонстрировано применение генетического алгоритма к многокритериальной задаче оптимизации учебного плана.

Ключевые слова — компетенции, учебный план, ориентированный граф, генетический алгоритм, методы оптимизации.

Список лит.: 12 назв.

UDC 004. 02:378

Application of Genetic Algorithm for Curriculum Optimization

Kurilova O. L. IUS, 2013. N 3. P. 84–92.

There has been presented an algorithm for optimization of curriculum in the framework of the competence approach. An algorithm for constructing adjacency matrix of subjects, an algorithm of a directed graph of subjects, an algorithm for finding the longest path in the graph have been developed. Application of genetic algorithm for multi-criteria optimization problem of curriculum has been shown.

Keywords — Competencies, Curriculum, Directed Graph, Genetic Algorithm, Optimization Techniques.

Refs: 12 titles.

УДК 004.932.2

О векторном квантовании изображений

Крук Е. А., Сергеев М. Б. Информационно-управляющие системы, 2013. № 3. С. 93–96.

Рассматриваются возможности использования кодов, исправляющих ошибки, в задачах векторного квантования для сжатия изображений. Предлагаются методы выбора кода и процедуры проведения преобразований изображения, позволяющие согласовать код с конкретным сжимаемым изображением. Обосновывается выполнимость процедуры сжатия с помощью декодера линейного кода. Показывается, что при кодовом квантовании достигается требуемый уровень сжатия при меньшей сложности (за меньшее время), чем у известных методов векторного квантования.

Ключевые слова — сжатие изображений, векторное квантование, кодовое векторное квантование, кодовая книга.

Список лит.: 11 назв.

UDC 004.932.2

Vector Quantization of Images

Kruk E. A., Sergeev M. B. IUS, 2013. N 3. P. 93–96.

There have been considered opportunities of using codes which correct mistakes in vector quantization problems for image compression. Methods of code selection and image transformation procedures allowing coordination of a code with a certain compressed image have been proposed. It has been shown that code quantization provides achievement of a required level of compression more easily (in a shorter period of time) compared to the known methods of vector quantization.

Keywords — Image Compression, Vector Quantization, Code Vector Quantization, Codebook.

Refs: 11 titles.